

**DATABASE CREATION, MANAGEMENT, AND INTEGRATION NOVEL  
METHODOLOGIES, TECHNIQUES AND TECHNOLOGIES**

**Final Report**

**Submitted to**  
**The U.S. Department of Energy**

**DOE Patent Clearance Granted**

*Mark P. Dvorscak*

*July 5 2005*  
**Date**

**Mark P. Dvorscak**  
(630) 252-2393

**By**

E-mail: mark.dvorscak@ch.doe.gov  
Office of Intellectual Property Law  
DOE Chicago Operations Office

**Jackson State University  
HPCC Research and Training Office  
P. O. Box 18180  
Jackson State University  
Jackson, Mississippi 39217-0480  
Tel.: 601-979-3303 FAX: 601-979-3310**

**Participating Science and Engineering Alliance Institutions**

**Alabama A&M University  
Jackson State University  
Prairie View A&M University  
Science and Engineering Alliance  
Southern University and A&M College**

**Project Director: Kunal Ghosh**

**The U.S. Department of Energy Grant Award No. DE-FG05-94ER25229**

## **DISCLAIMER**

**This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.**

## **DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

## **Section 1. Scientific Database Management System**

**H. John Caulfield**

**Marius Schamchula**

# OPTICAL DATABASE MANAGEMENT SYSTEM FOR POHM's (PAGE ORIENTED HOLOGRAPHIC MEMORIES)

H. John Caulfield  
Marius Schamachula  
Jian Fu  
Rajani Srinivasan

## Objective

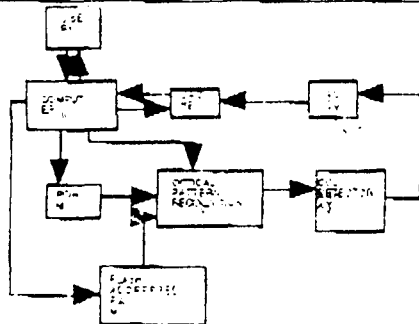
It is now widely recognized that POHMs can have good storage capacity, say,  $10^{15}$  bits and superb random access rates say  $10^{12}$  bits/second or better. This is accomplished by storing  $10^3$  or more "pages" can be randomly accessed at  $10^6$  or more per second. The trouble, of course, is that we do not know how to handle  $10^6$  bits in parallel every microsecond using electronics. Conclusion: the data use rate is slowed down by the electronics. We would not only have to read  $10^{12}$  bits per second but also perform complex operations across a field of  $10^6$  bits each microsecond. Solution: Process the parallel optical pages of data in parallel using optics.

For our initial studies, we have lumped object-oriented and relational databases together as one topic and consider image databases as a second topic.

## Methods

We have tried to outline and explore generic systems for both cases. The systems includes hardware, software, and interconnections. Then we devised specialized systems for both cases

## GENERIC SYSTEM DESIGN



Moving now to image databases, we encounter more difficulty. Different images of "the same thing" may look very different. Certainly, size, pose, lighting, etc. are likely to vary. How can we search for patterns in such a case?

Optical syntactic pattern recognition is a highly robust way of recognizing some simple objects. The basic idea of this approach is to recognize the presence, location, and sometimes, orientation of a set of features by separate Fourier filtering operations. We then combine those data into some sort of "score" which we use to make a decision. We can tune the scoring to get substantial agreement with human performances or predispositions.

Consider the letter A as being comprised of three primitives or sub-patterns,  $a_1$ ,  $a_2$ , and  $a_3$ , as shown in Fig. 1. The scene consisting of the letter A can be described as follows: There is an  $a_1$ -like feature above and to the right of an  $a_2$ -like feature, as well as above and to the left of an  $a_3$ -like feature. And there is an  $a_2$ -like feature below and to the left of an  $a_1$ -like feature, as well as to the left of an  $a_3$ -like feature. The line through  $a_2$  and  $a_3$  is more or less horizontal. Similarly, there is an  $a_3$ -like feature below and to the right of an  $a_1$ -like feature, as well as to the right of an  $a_2$ -like feature. This is a fuzzy description and the scoring procedure provides quantitative evaluations of terms such as  $a_1$ -like, and more or less.

Matched filters of the primitives were correlated with input characters, which were sets of computer generated and handprinted A's. The handprinted characters were all roughly the same size. Ideally, correlation of a perfect A with each of the researched primitives should be characterized by

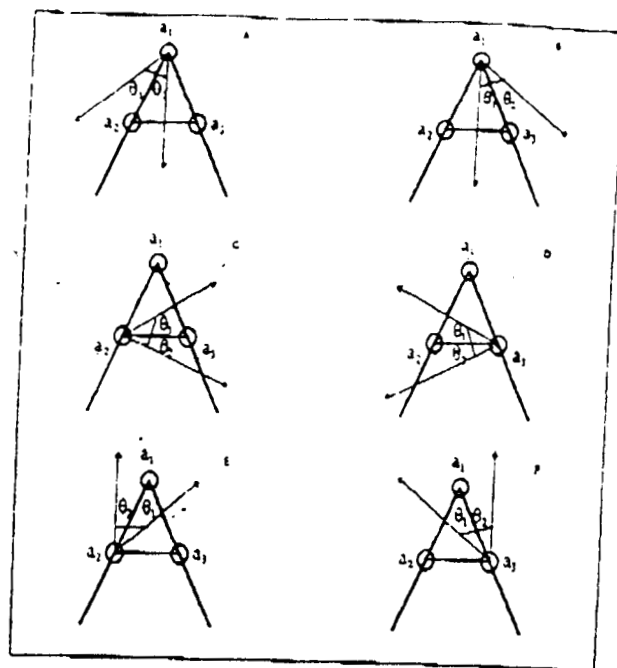
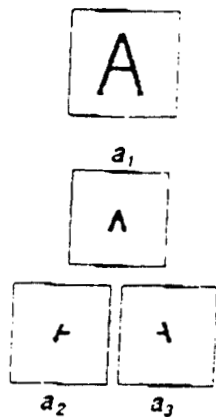


Fig. 2. Fans drawn from correlation point  $(x_1, y_1)$  on correlation surface  $a_1$  to encompass correlation point  $(x_1, y_1)$  on correlation surface  $a_2$ . The circles represent the primitive correlation surfaces. The fuzzy membership values are a maximum along the bisector of the total fan angle for a, b, e, and f, and along the horizontal axis for c and d.

Table 1. Normalized scores obtained for some test characters.

Normalized Score	A1 (computer generated)	A2 (computer generated)	A3 (Hand printed)	A4 (Hand printed)	A5 (Hand printed)	A6 (Hand printed)
	A	A	A	A	A	A
$N_1$	1	1.4	0.96	.67	0.009	.36
$N_2$	1	2.48	1.009	.66	0	0
$N_3$	1	2.42	1.21	.76	0	0
$N_4$	1	2.57	0.77	.39	0	0

Physically, it is the significance of the location of feature  $j$  at  $(x_j, y_j)$  in view of the assumed presence of the feature  $i$  at  $(x_i, y_i)$  in confirming the nature of the object. For a detailed description of the fans and determination of fuzzy membership values.

Using the correlation peak values and the fuzzy memberships, we have tried four possible scoring methods which are:

$$N_1 = \sum_i \sum_j \mu_{ij} S_j, \quad i \neq j. \quad (2)$$

$$N_2 = \sum_i \prod_j \mu_{ij} S_j, \quad i \neq j. \quad (3)$$

$$N_3 = \sum_j \prod_i \mu_{ij} S_j, \quad i \neq j. \quad (4)$$

$$N_4 = \sum_i \prod_j \mu_{ij} S_i S_j, \quad i \neq j. \quad (5)$$

For this study  $i, j = 1, 2, 3$ .

We concluded that the scoring method " $N_4$ " conforms closest to our intuition as it uses multiplicative features which are high only if all features are present and well-formed.



one sharp peak in the image plane [1]. The maximum normalized peak values ought to be  $\pm 1$ . This would not be the case for real A's, where the correlation peaks could be small.

For scoring each pattern with respect to our description of A, the data that are required are the normalized correlation peak heights  $S_i(x, y)$  and their location points  $(x_i, y_i)$ . The scoring method involves setting up a "fuzzy fan" about each correlation point  $(x_i, y_i)$  on the correlation surface  $a_i$  of the pattern. The fans are extended to encompass every other correlation point  $(x_j, y_j)$ , where  $i \neq j$ , as shown in Fig. 2. The points  $(x_j, y_j)$  on the correlation surfaces of the primitives could lie anywhere between the fan boundaries, or outside the fans specified, depending on the type of input character. Since three primitives were chosen for the character A we have a total of six fans, two from each correlation point.

We define  $m_{ij}$  as the fuzzy membership of the feature in the class of primitive  $a_j$ , in the cone enclosed by fans drawn from the primitive correlation point  $(x_i, y_i)$  on correlation surface  $a_i$ . For six fans we have six fuzzy membership values  $m_{ij}$ ,  $i, j = 1, 2, 3$ , where  $i \neq j$ .  $m_{ij}$  is a function of the fan angle  $\theta$ , which is determined by the correlation point  $(x_j, y_j)$ . Mathematically we can express  $m_{ij}$  as:

$$\mu_{ij} = f(\theta) = f(x_j, y_j | x_i, y_i). \quad (1)$$

- Research

- Generic plans/ architecture
- Generic encryption
- Correlator Design
- Optical Fuzzy Syntactic  
Pattern Recognition

- Student Involvement

- Jian Fu (MS 1996)
- R. Srinivasan (projected Ph.D. 1996)

- Journal Articles

## REFERRED PAPERS

- Published Papers

R. Srinivasan, J. Kinser, M. Schamschula, Joseph Shamir, and John Caulfield,  
"Optical Syntactic Pattern Recognition by Fuzzy Scoring," Optics Letters **21**, 815 (1996)

- Invited Papers for Pattern Recognition R. Srinivasan, H. John Caulfield, and Alex Granik, "Perfect and Near-Perfect Pattern Recognition," Pattern Recognition, Invited Papers, Accepted (1996)

- In Preparation

H. John Caulfield, M. Schamschula, Jian Fu, "Optical Parallel Database Management System for One Page Holographic Memories"

- Submitted to APPLIED OPTICS

H. John Caulfield, M. Schamschula, Jian Fu, "System Design of Database Management for One Page Holographic Memories"

## Presentations

- Marius Schamschula
- John Caulfield (SPIE)

- Collaborators to Date

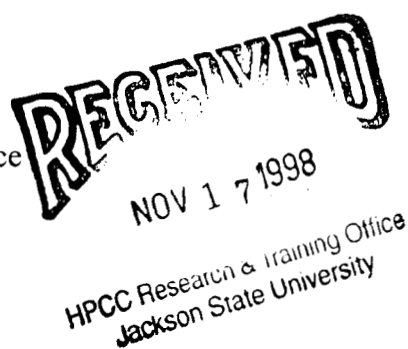
- Joseph Shamir - The Technion
- Alex Granik - University of the Pacific

**Section 2. Integrated Scientific Simulation Environment Based  
on Distributed Computing**  
**Gwang S. Jung**  
**Jackson State University**

## SEA-DOE Project Report

Gwang S. Jung  
Department of Computer Science  
Jackson State University  
Jackson, MS 39217

November 17, 1998



### 1. Brief Background About the Research Grant Award

The Science and Engineering (SEA) is a collaboration, formed in 1990, among five institutions: Alabama A&M (AA&M), Jackson State University (JSU), Southern University at Baton Rouge (SUBR), and Prairie View A&M University (PVA&M), and Lawrence Livermore National Laboratory. The SEA partnership was established to encourage and increase participation of faculty and students in the four HBCUs in high quality scientific research. The design team for the High Performance Computing and Communication (HPCC) initiative from the five component institutions of the SEA met during the spring of 1993 to prepare a Talking Paper for Marketing Purposes (TPMP) based on the expertise and the needs available in each institution. the TPMP had an overview of the different research efforts of the four SEA universities in the general areas of the HPCC.

After review from DOE, four projects have been selected for full proposal preparation. In April, 1994, a proposal entitled "Database Creation, Management and Integration: Novel Methodologies, Techniques and Technologies" was submitted to the DOE. Six PIs were involved in this project: Dr. Kunal Ghosh in the Department of Physics and Atmospheric Sciences at JSU as the Project Director, Dr. John Caulfield of the Physics Department at AA&M, Dr. Jiang D. Fan of Physics Department at SUBR, **Dr. Gwang S. Jung** of the Department of Computer Science at JSU, and Dr. Anil Kumar in the Department of Electrical Engineering at PVA&M. Dr. Robert Shepard of SEA Executive Office was involved as an overall project coordinator. The project was funded by DOE and the amount awarded was in total \$1,080,000 for the five PIs and one coordinator for three calendar years (final revised funding period: September, 1994 to August, 1997).

### 2. The Project Performed by Dr. Gwang S. Jung

The amount awarded for Dr. Gwang S. Jung's research project was \$210,000 for three calendar years (grant number: DE-FG05-94ER25229). Throughout the project, the PI could be released from one course teaching load, partial summer support, one graduate student. One undergraduate student is supported for 6 months.

Through the research project, the PI could establish research infrastructure including computer workstations dedicated for conducting DOE research. The DOE research fund significantly improved research infrastructure. This fund was greatly helpful for the PI to develop research capability in the research areas of distributed multimedia information storage and retrieval.

## 2.1. Research Objectives

The main research objective of the proposed research was investigating various research issues for building an Integrated Scientific Information Service Environment which provides the scientists with the following services.

1. Locating and retrieving information about the experimental data and resources.
2. Locating and retrieving information about software artifacts (software codes, designs, simulation packages).
3. Locating and retrieving relevant information for the scientist's research work.
4. Providing Database storage and retrieval facilities for storing retrieved data/ information, and experimental results.

## 2.2. Research Accomplished

As the main objective represents, the major research effort for the project was devoted for investigating research issues through a research prototype design and implementation. A research prototype, named JSU-IIS (Integrated Information Server) has been designed and *partly* implemented for providing the services enumerated above. IIS consists of information retrieval manager (IR Manager) and Information Search Agent (ISA). The overall architecture of the JSU-IIS is shown in the Figure 1.

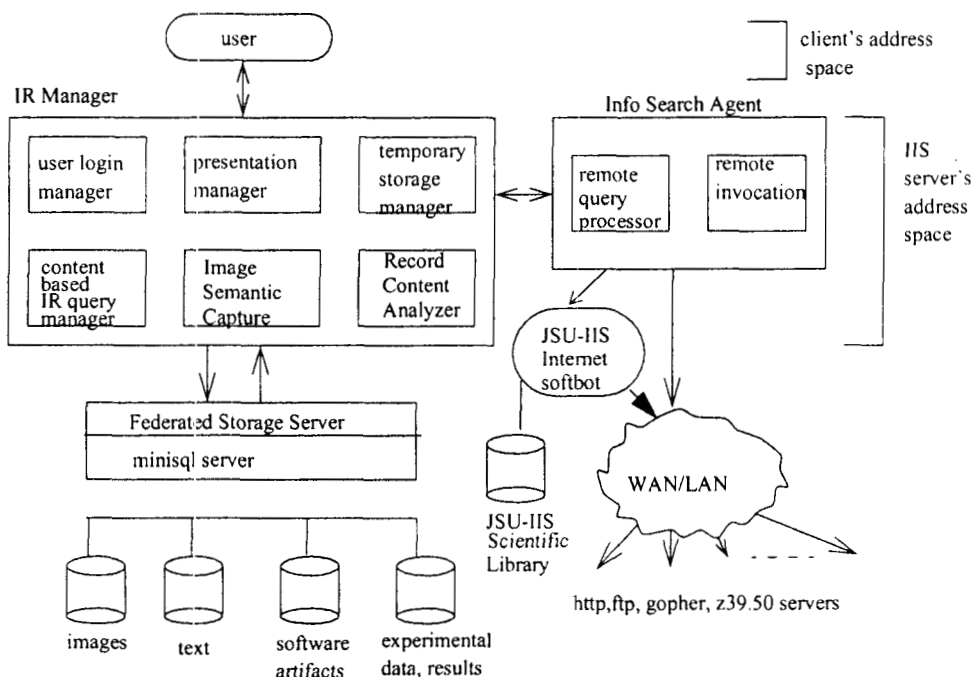
IR manager provides facilities by which the end user can store and retrieve data and information for their research. A user can retrieve data and information from Internet sites through the service provided by the ISA. A user can seamlessly structure a query using the GUI provided by IR manager. If the user wants to store the downloaded data and information for future reference, the user can ask the IR manager to store them in his/her database tables. The database tables for different users are managed by the IR manager communicating with the minisql DBMS server. Minisql server is used for storing data such as users' login information, structured data such as book/journal information.

Information Search Agent is responsible for sending queries to about 100 predetermined Internet information sites (mainly HTTP and Z39.50 servers) and JSU-IIS Scientific Library. Some of the predetermined sites include software archive sites, scientific data repository sites, and Virtual Web Libraries.

JSU-IIS scientific virtual library is a small-scale Internet Library in Physics, Chemistry, and Biology Subjects. It contains 1,000 HTML documents in this field of research. Each document consists of a hyperlink to a remote HTML document (or a Web site) and the abstract of the HTML document. The content of the our virtual library is periodically updated by new documents gathered by JSU-IIS Internet softbot.

The major components of the JSU-Internet softbot are navigator, indexer, URL database. The navigator is responsible for getting newly updated information about scientific libraries including about 200 Virtual Web Libraries in Physics and Chemistry domains. The navigator automatically gathers useful documents by fine-grained content analysis of the documents. To gather other useful documents beyond the Virtual Web Libraries, the gatherer uses three other information search tools including Lycos, InfoSeek, and Altavista.

The main function of the gatherer is collecting URLs of the documents that are potentially relevant to the scientific categories. The indexer then perform automatic indexing based on modified *tf\*idf* indexing scheme to build inverted URL database. Each entry in the URL base is an ordered pair of URL and several keywords extracted from the document corresponding to the URL.



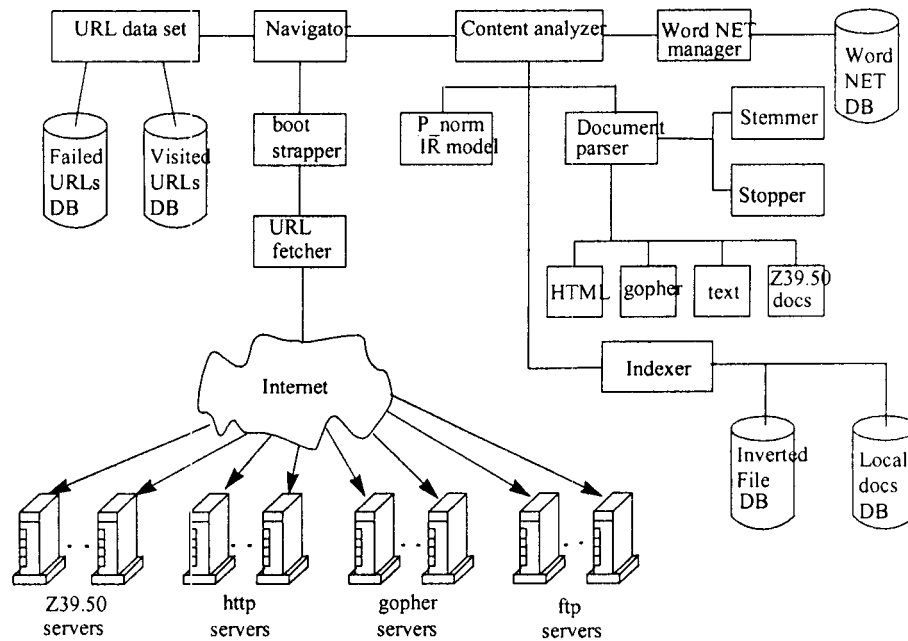
**Figure 1. Overall Architecture of JSU-IIS (Integrated Info Server).**

Through the design of the JSU-IIS, theoretical and technological research issues have been investigated. The research investigations have been reported and published in the following papers (Copies of the papers are enclosed in the packet).

- "An Architecture for and Query Processing in Distributed Content-based Image Retrieval", *Real-Time Imaging Journal*, Academic Press, Vol. 2, 1996, pp. 139-152. (with V. Gudivada)
- "Distributed Adaptive Attribute-Based Image Retrieval", in the Proceedings of the 1st International SPIE Conference on Image Storage and Archiving Systems, Vol. 2606, SPIE Press, 1995 (with V. Gudivada)
- "Autonomous Tools for Information Discovery in the World Wide Web", *Technical Report*, Department of Computer Science, Jackson State University, August, 1995.

Based on the research experience obtained from the project, the PI could submit research proposal titled "*Enabling Technology for Building High Performance Distributed Information Storage and Retrieval System*". The research proposal was selected to be funded

by the DOE for the period from September 1997 to August 1999



**Figure 2. The Architecture of JSU-IIS Internet Softbot.**

**Section 4. Data Management and Database Creation**  
**Anil Kumar**  
**Prairie View A&M University**



# DATA MANAGEMENT AND DATABASE CREATION

**Final Report to the  
Department of Energy**

*for the period*

15 September 1997 – 14 September 1998

**Grant #DOE-DE-FG05-94ER25229**



**A. Anil Kumar, Ph.D.**

Director of Research and Special Assistant  
to the President for Science & Technology

*Office of Research & Development*

*Building 1, Administration Building,*

*Room 108 L. W. Minor St.*

*Prairie View, TX 77446-4149*

*Tel: 409-857-2591, Fax: 409-857-2255*

*[Anil\\_Kumar@pvamu.edu](mailto:Anil_Kumar@pvamu.edu)*

**Adriana Caceres, M.S.E.**

Research Associate &

Visiting Instructor

*Electrical Engineering Department*

*College of Engineering*

*Ann Preston & O'Banion St.*

*Prairie View, TX 77446-0397*

*Tel: 409-857-4518, Fax: 409-857-4780*

*[yalily@ieee.org](mailto:yalily@ieee.org)*

**PRAIRIE VIEW A&M UNIVERSITY**

# DATA MANAGEMENT AND DATABASE CREATION

A. Anil Kumar, Principal Investigator  
Adriana Caceres, Research Associate  
*Electrical Engineering Department*  
*Prairie View A&M University*

## I. OBJECTIVES

The major objectives of this project are:

- to enhance the existing infrastructure at Prairie View A&M University in the general area of High Performance Computing and Communications (HPCC), in particular to address problems related to DOE's Grand Challenges;
- to establish a strong research group with sufficient hardware and software needed to:
  - test and evaluate available software for use in performing high performance simulations in science and engineering, and
  - investigate systematically existing and planned trends in high performance computing architectures and provide predictions on future trends,
- to identify partners from industry and government laboratories, and establish potential collaborative interactions and additional funding sources; and
- to educate and train students for advanced careers in the HPCC arena.

## II. OVERALL SCOPE OF WORK

The work originally planned was to focus mainly on developing a relational object oriented database model for specific scientific and engineering applications, such as materials simulation, which is one of the grand challenges of the HPCC Program and in particular of the Department of Energy [1-5]. The motivation was that while the processor unit speeds and costs have attained newer limits (owing to new logic technologies as well as to the availability of parallel processing), the software component was largely ignored in that it had not kept up pace [6-10]. Even here, emphasis was more on speed and numerical accuracy rather than software design practices that would allow adaptability and reuse for other similar calculations. The extra time and effort devoted to such design practices will be repaid in the reduced time required for testing and debugging. These issues will be particularly important in view of the expected scalability and interoperability of the future large-scale distributed and heterogeneous computing environments. In addition, it also has been realized that hardware and software improvements need to occur with the *overall* performance of the system in sight.

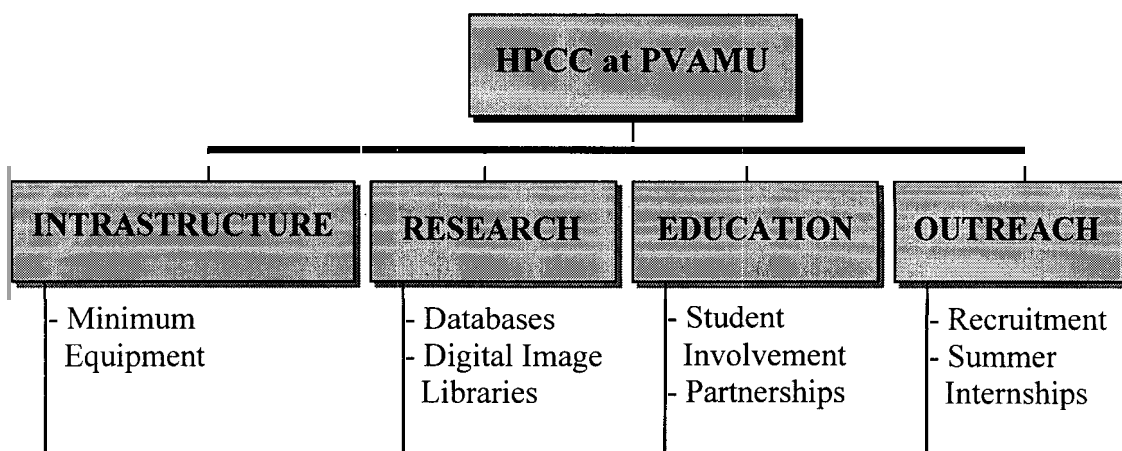
The specific scope of the work was to develop an infrastructure in the HPCC area, develop new algorithms for science and engineering database management, employ commercial data management systems and develop human resources in the HPCC area. This research will allow commercialization and technology transfer to DOE, its national laboratories and the industry. This goal has been achieved and several details were provided in the Third Annual report,

September 1997. Further research was conducted after obtaining permission from DOE to spend remaining funds. The two major topics to address were image storage algorithms employing wavelet transforms and database design for the internet. The first objective has been achieved with tremendous success. Indeed, three major proposals applying this newly established expertise to different applications were prepared. Also two extended abstracts were submitted and accepted at a national conference (see Section ???). The internet database design however was completed partially owing to the graduate research assistant leaving for an industrial employment.

The major outcomes of this effort are:

- **preliminary prototypes of an in-house developed algorithm based on wavelet analysis for image compression and decompression,**
- **identification of several “off-shoot” areas of HPCC research highly relevant to DOE (in particular DOE2000),**
- **two completed Master’s theses,**
- **training of undergraduate and graduate students in the energy research areas, and**
- **the incorporation of these concepts into the curriculum.**

The current proposed project addresses the issue of wavelet characterization for image libraries, and an iterative process of image compression looking forward to find the “fixed point” to store a compressed image in the minimum memory space as possible. One of the goals of this research is to define the parameters involved in the iterative compression scheme - number of levels of decomposition, optimum threshold value, and number of iterations. These results are of immense importance for applications to image databases for the next generation Internet, tele-medicine, and telecommunications, to name just three examples.



**Figure 1 Overall activities during the period of this grant**

### III. SCOPE OF WORK - 1997 - 1998

The specific topics addressed this year are image processing using an iterative wavelet approach and database design in internet applications. These subjects are continuation of the thesis completed last year – “Image Processing Techniques for Digital Image Libraries” by *Mayra Caceres*. The following areas are an introduction to the relevance of these two topics.

#### ***Image Processing***

Targeting the Global Information Infrastructure, it is important to establish the foundation for information systems capable of linking, discovering and utilizing diverse computational resources and data sets in an intelligent scaleable manner. Image processing – compression, storage, and retrieval – is a major element in this infrastructure. In relation to this work, a standard color image with 640 x 480 pixels and 24 bits resolution allowing for 256 shades in each color component requires approximately 1 MByte. A video-clip needs more than 22 MBytes of storage. Quite apart from the storage requirements this data must also be transferred from a disk or CD-ROM into the processor memory before it can be displayed. In the case of distributed multimedia systems this data must be communicated over the network.

The main difficulties associated with image compression methods are:

- The compression - and often also the decompression - process is usually very time consuming. This fact usually makes real-time compression and decompression unfeasible, except if specially designed hardware is used.
- Depending on the compression method used, the quality in the decompressed image may be unacceptable for certain applications.
- Some compression methods are sensitive to noise in the original images, such images obtained with medical ultrasound equipment. In this case more processing is required to enhance or restore the image quality.
- Standardization of compression methods is essential if these are to be used on a universal scale.

#### ***Databases***

The databases designed for the internet have several shortcomings, especially when scientific and engineering databases are concerned. The design of these databases requires a lot more precision and detail than generic commercial databases. The tenet behind this project is that design of databases has to be done with the design of the whole system including the search engine(s).

### IV. ACCOMPLISHMENTS & RESULTS

The completed results are:

- a novel multi-stage, multi-resolution algorithm employing iterative wavelet based techniques,
- results based on database performance in relation to searching and indexing.
- two abstracts approved to be published in ANNIE '98 Smart Engineering System Design, International Conference.
- three papers in preparation.
- two theses published in the areas of image processing and channel characterization.

The two theses are available as separate reports on request. The summaries of the theses' results are included in Appendix B.

The following is a brief summary of results obtained in each of these areas.

#### **IV.1 IMAGE PROCESSING ANALYSIS WITH ITERATIVE WAVELET APPROACH**

**Adriana Caceres & A. Anil Kumar**

A major motivation for wavelet based image processing is the fact that in addition to achieving vast savings on memory space it is also possible to employ an iterative process (discussed below) of transforming the original image into images of lower resolution through successive stages of wavelet analysis. The advantage of this process is that storing all the images at various stages (low resolution images - LRIs) requires at most the same memory as that of the original image (where such fidelity is required) and hence this process is more memory efficient. In the standard approaches however, a number of images at various resolutions need to be stored, say for browsing purposes, which is not memory efficient. Also, in wavelet based schemes, the LRIs with increasing resolutions can be stored in progressively slower memories, so that low resolution images may be accessed faster. This would be very relevant for browsing through the database or the image library.

While an enormous amount of mathematical literature exists on the subject of wavelets, the explicit implementation in a given application poses several questions. One such question is in terms of the choice of wavelets. There are a vast number of possible choices for the wavelet analysis. The following facts are generally known:

- the wavelet filters should be small so as to keep the computational burden low;
- the regularity of the wavelet basis functions has an important role in removing blocking artifacts;
- some wavelets, such as Daubecheis wavelets, with longer lengths tend to yield over smoothed images with lower perceptual quality.

However, explicit implementation does require a vast amount of knowledge base as well as appropriate benchmarks.

To achieve compression using wavelets we follow the following procedure:

- Load the image to be compressed
- Select the wavelet family and the specific wavelet to be used for the analysis
- Perform discrete wavelet transform (DWT) to the image to get the image decomposition.
- Analyze the coefficients obtained from the decomposition and discard as many coefficients to keep as much information of the image as possible. In other words, trying to maintain the distortion at the minimum level. At this point we have the compressed image.
- Reconstruct the image from the compressed image performing inverse DWT. This is to ensure that the quality of the reconstructed image is acceptable by the application.

Let us call all the process just described as the Wavelet Compression Algorithm (WCA). By optimizing the selection of the threshold we can achieve the highest compression ratio with the lowest distortion of the reconstructed image.

### **Image Quality Measurements**

No single approach to quality measurement has gained universal acceptance. However, three general approaches have come to dominate:

1. computable objective distortion measures (Mean Squared Error (MSE) or Signal-to-Noise Ratio (SNR)),
2. subjective quality (psycho-physical tests or questionnaires with numerical ratings), and
3. simulation and statistical analysis of a specific application of the images, (diagnostic accuracy in medical images measured by clinical simulation and statistical analysis).

For medium bit rate (0.25 bpp), the objective measures are good indicators of the subjective quality of the image. At low bit rate, this is not always the case. There are many artifacts in the reconstructed image: the choice of filters is important as well as bit allocation.

Bit allocation plays an important role in decreasing the distortion effects mentioned before. The blurring effect will decrease if one allocates more bits to the higher subbands (wavelet compression).

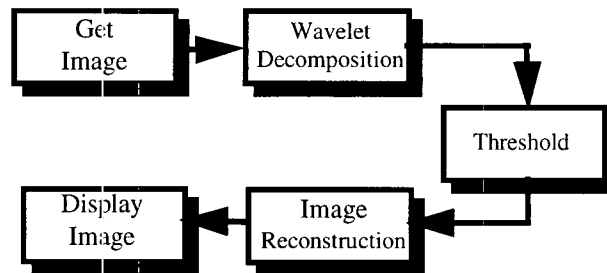
### **Iterative Procedure**

The previous discussion describe the procedure to compress gray-scale images however, several issues related with different steps in the compression techniques need to be considered:

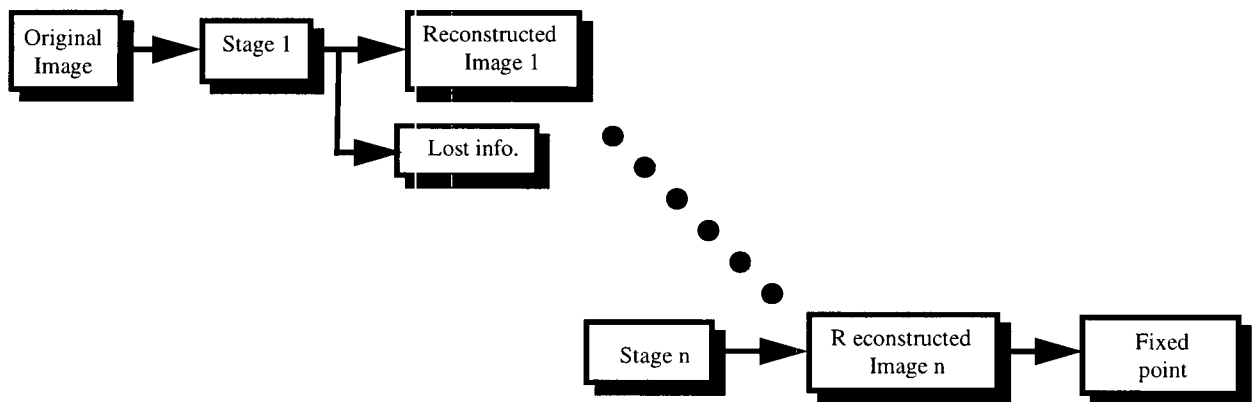
- First, when loading the image we have to know or select the format we are using to represent that image. For B&W images, the intensity of each pixel is giving at different resolutions depending on the application, with typical values of 8 bpp (bits per pixel) giving 256 shades of gray. For color images there are various formats that may introduce some differences to the reconstructed image, even using the same general procedure to compress it.
- Second, the selection of the wavelet is a very important issue and there is not a standard or optimal technique available for this.
- Third, the steps within a given stage of image decomposition can have as many levels as desired, but from experience we know that 4 to 5 are usually adequate.
- Fourth, discarding coefficients according to certain criteria. This is a very important step and the final distortion of the reconstructed image depends on this.
- A threshold value needs to be considered such that for each level of decomposition some of the coefficients containing the information of the horizontal details, vertical details and diagonal details of the image will be discarded.

This new approach consists of running iteratively WCA such that we can improve the results in terms of higher compression ratios with less distortion. In this case we have to define the parameters (number of levels of decomposition, threshold value and number of iterations) for the basic WCA, where we try to optimize only in terms of low distortion. Then the reconstructed

image obtained after running the basic WCA will be used as an input to run WCA again without modifying any parameter. The number of iterations will be determined when the “fixed point” is achieved, viz., the compression ratio does not increase any more with more iterations of this process. Figures 2 shows a schematic of the single stage compression for the WCA, and Figure 3 shows the proposed iterative multi-stage compression process.



**Figure 2 - MATLAB implementation of Wavelet-based compression algorithm**



**Figure 3 – Wavelet Iterative Processing**

### Color processing

The processing for color image follows the same procedure but some modifications need to be done due to the different formats to handle the image. A short summary of formats is as follows.

The most common formats used are:

- **Indexed images:** It requires two matrices to represent the image. One matrix has as many elements as number of pixels the image has. Each element indicates the index of the color the correspondent pixel has. The other matrix is the colormap, that is the definition of each color used in the image.
- **RGB images:** The image is represented by three matrices. Each of this having as many elements as number of pixels the image has. The color of each pixel is obtained by looking up at the three matrices R, G, and B and combining the red (R), green (G) and blue (B) components. Compressing the image means to run the compression procedure for the three matrices.

- **Y-Cr images.** The image can be represented by the luminance and Chrominance components. Where it is basically the same procedure in the sense that both matrices have to be processed and some considerations have to be taken into account.

A picture “barbara” was analyzed using this iterative technique. The original image and the “fixed point” images are shown in Appendix A.

## **IV.2 DATABASES - SEARCH ENGINES**

### **Gerardo Novelo, Singh Madden & A. Anil Kumar**

#### **TERMS AND DEFINITIONS**

##### **Evaluation Methods**

- Relevant Sites – Sites that have useful information that pertained to the keyword
- Irrelevant Sites- Sites that may contained the keyword but the information is misleading or “junk”.
- Not Accessible – Sites that have the information in different format (not readable) or the web page is no longer on the Internet.

##### **Search Methods**

- Word Combination – The search engines browse the Internet to find websites (articles, publications, etc) that include the keyword(s) in its title in any format.  
Example: Air versus bone conduction: An equal loudness investigation.  
The Bone Conduction Show
- Any Word Match – The search engines execute the query by matching any word from the keyword(s) with the web site title.  
Example: Bone marrow donors.  
Heat conduction.
- Intelligent default – The search engines execute the query by matching the keyword(s) with related titles.  
Example: Bone Anchored Hearing aids.  
Subliminal use of bone conduction.  
Audiologist hearing evaluations using air and bone conduction.
- Exact phrase match – The search engines execute the query by exactly matching the keyword(s) with the titles.  
Example: The use of Bone Conduction hearing aids.  
The effect of preoperative bone conduction threshold levels on the results of surgery.

#### **DATA EVALUATIONS**

The data in table 1 shows the efficiency response of three search engines with respect to a specific query using the keyword Bone Conduction. Each search engine produced a countless number of possible matches, but for this evaluation, only the first 100 possible matches were considered. The evaluation chart consisted of three categories Relevant Sites, Irrelevant Sites, and Not Accessible. Altavista proved to be the best search engine in comparison with Excite and



Yahoo. It had a greater number of possible matches, a higher percentage of relevant sites, and a lower percentage in irrelevant sites and not accessible categories. The search engine Yahoo offered the option to refine the query by selecting other search methods like “exact phrase match”. In this case the number of possible matches was reduced significantly when compared to the search method “intelligent default”. In this comparison “exact phrase match produced a 5% more relevant sites and 4% less not accessible sites than its counter part “intelligent default.

**Table 1. Efficiency of Internet Search Engines**

Keyword used for search: *Bone Conduction*

	Alta Vista	Excite	Yahoo	*Yahoo
Possible Matches	307600	173117	195815	297
Relevant Sites	61 %	19 %	26 %	31 %
Irrelevant Sites	35 %	74 %	68 %	67 %
Not Accessible	4 %	7 %	6 %	2 %
Search Method	Word Combination	Any word match	Intelligent default	Exact phrase match

\*Refined Search

The data shown in table 2 was the result of a search engine called “ProFusion”. ProFusion sends the query to multiple search engines at the same time. Search results returned by all search engines are then combined with duplicates removed and relevance factors recalculated. Therefore, you can take advantage of multiple search engines and obtain documents that are more likely to meet your expectations. In this case six different search engines were evaluated. Infoseek produced more possible matches with more relevant sites and less irrelevant sites. Yahoo was not as efficient because it did not produce any results.

**Table 2. Efficiency of ProFusion Meta Search Engine**

Keyword used for search: *Bone Conduction*

	Alta Vista	Excite	Yahoo	Web Crawler	Magellan	Infoseek
Possible Matches	10	10	0	25	10	18
Relevant Sites	50 %	20 %	0 %	8 %	10 %	67%
Irrelevant Sites	40 %	60 %	0 %	92 %	90 %	28 %
Not Accessible	10 %	20 %	0 %	0 %	0 %	5 %

Table 3 and 4 illustrates results produced by the similar search engines as in table 1, but with different word format and additional selected words for the keyword. When that keyword was place within quotation marks (“bone conduction”), the number of possible matches was significantly reduced for each search engine. In this case Altavista was not the most efficient search engine as in table 1 but rather Excite. When compared to Altavista, Excite had 5% more relevant sites and 7% less not accessible sites but 2% more in irrelevant sites. In table 4 additional select words were introduces to the keyword but in this case the results were not favorable. In comparison to table 1,2,3 the number of possible matches was greater for the three search engines and the efficiencies were reduced significantly. These unfavorable results may have been generated due to the fact that the keyword contained more words therefore the search engines responded accordingly.

**Table 3. Using common “denominator” for each Search Engine**Keyword used for search: “*Bone Conduction*”

	<b>Alta Vista</b>	<b>Excite</b>	<b>Yahoo</b>
Possible Matches	898	282	297
Relevant Sites	28 %	33 %	23 %
Irrelevant Sites	63 %	65 %	69 %
Not Accessible	9 %	2 %	8 %

The results indicate that the search engines are efficient depending on the format of the keyword(s) used for the query. In general Altavista was the most efficient search engine to execute queries while Excite and Yahoo came in correspondingly.

**Table 4. Refined Search using common “denominator” and additional selected words**Keyword used for search: “*Bone Conduction*” *hearing, audiology, noise*

	<b>Alta Vista</b>	<b>Excite</b>	<b>Yahoo</b>
Possible Matches	556505	281560	195815
Relevant Sites	9 %	14 %	16 %
Irrelevant Sites	88 %	77 %	81 %
Not Accessible	3 %	9 %	3 %

**NOTE:** All data is based on the first 100 possible matches

## V. PUBLICATIONS AND PRESENTATIONS

Two abstracts approved to be published for ANNIE 98 Smart Engineering System Design, international conference.

1. Iterative Wavelet-based technique for Image Compression, Storage, Retrieval and Transmission, by A. Caceres & A. A. Kumar.
2. Noise Effects in Transmitting Compressed Information, by A. Caceres & A. A. Kumar.

Three papers in preparation:

1. *A Comparative Review of Modern Search Engines on the Internet*, by Gwang Jung & A. A. Kumar.
2. *Image Storage With Wavelets – Objective vs. Perceptive Criteria*, by A. A. Kumar & A. Caceres
3. *Increased Data Access Accuracy via a Successive Three-Stage Process*, A. A. Kumar & K. Paick.

## VI. PERSONNEL INVOLVEMENT IN THE PROJECT

Apart from A. A. Kumar, Principal Investigator, the senior member involved in the project is Adriana Caceres, Electrical Engineer. Her work is an extension of the Thesis "Image Processing Techniques for Digital Image Libraries" completed in May '97 by Mayra Caceres.

The students involvement during this year include one graduate (\*\* Hispanic) and one undergraduate student (\* African American). Table 5 shows their activities.

**Table 5 . Student involvement and activities.**

<b>Name/Major</b>	<b>Title</b>	<b>Task</b>	<b>Completed</b>
Mr. Gerardo Novelo (EE)**	Graduate Research Assistant	Research on Internet Search Engines	Dec. '98
Mr. Byron Page (EE)*	Undergraduate Research Assistant	Run Wavelet iterative simulations. Research on Wavelet Packets and MATLAB	May '99

## VII. ALLIANCES / COLLABORATIONS

Strong collaborations and future partnerships were developed during this period with some companies working in this field: Houston Advanced Research Center (HARC), American Heuristics Corporation (AHC), Compression Engines LLC and the Center for Research in Parallel Computation (CRPC) at Rice University. Table 6 contains their respective addresses.

**Table 6. Points of Contact**

<b>Agency</b>	<b>Address</b>
<b>Houston Advanced Research Center</b> David Norton	4800 Research Forest Drive The Woodlands, TX 77381 Tel: 713-363-7941, Fax: 713-363-7914 norton@harc.edu
<b>American Heuristics Corporation</b> Eric A. King	2 Millennium Centre Triadelphia, WV 26059 Tel: 304-547-4201, Fax: 305-547-4203 Eking@heuristics.com
<b>Compression Engines LLC</b> Jianrong Wang	16800 Greenspoint Park Drive, Suite 240 N Houston, TX 77060 Tel: 713-876-3966, Fax: 713-876-3974 Wang@cengines.com
<b>Center for Research in Parallel Computation</b> Ken Kennedy	Rice University 6100 Main St – MS 41 Houston, TX 77005-1892 Tel: 713-527-6009, Fax: 713-285-5136 ken@rice.edu

## VIII. IMPACT OF THE GRANT ON PVAMU

In addition to education and training of several undergraduate and graduate students, three proposals were submitted for funding based upon the infrastructure and expertise developed.

1. *"Iterative Compression Schemes for Dat/Image Processing in a High Performance Computing Environment,"* A. A. Kumar & A. Caceres, to Department of Energy, May 1997.
2. *"Image Characterization and Identification Techniques – Applications to Biotechnology,"* A.A. Kumar, S. Shakir and A. Caceres, to the Texas Advanced Technology Program, July 1997.
3. *"Application of Non-Gaussian Signal Processing to Energy Related Problems,"* S.T. Koay & A.A. Kumar, to the Texas Advanced Technology Program, July 1997.

The following proposal, between Prairie View A&M University, Jackson State University and Tennessee State University, is under preparation.

*"Enabling Technologies for Digital Libraries - High Performance Multimedia Data Storage and Delivery in the Internet Environment,"* A. Anil Kumar & Kwang Paick (PVAMU), Gwang Jung & Q. Malluhi (JSU) and Paul Devgan (TSU)

*Brief Summary of Proposal:* This proposal outlines a comprehensive approach to digital libraries - creation, storage, access and transmission. Special consideration will be given to end-user search needs in digital libraries. It draws on a number of research projects underway at Prairie View A&M University, our collaborating institutions Jackson State University and Tennessee State University, as well as a review of the literature. The research discussed here is a part of a comprehensive proposal to the National Science Foundation's initiative on digital libraries. The major issue is the question of the creation of digital libraries, how users will be able to gain access to the contents of digital libraries, and of how information sources in a digital library can or should be identified, organized and retrieved to best support the users who need to have access to it. The answer to these questions lies in the relationships among information retrieval, organization, navigation and visualization, and should lead to attempts to determine if there is a common framework in which they can all be viewed.

The goal of this research is to develop a core enabling technology for building digital library, namely, high performance distributed multimedia storage and delivery system. The proposed method is based on stripping a large multimedia data (e.g., BLOB, or MPEG movie file, etc.) into blocks that are distributed across available data servers (e.g., Internet data servers, or Web servers, etc.). Efficient data service can then be achieved by using multiple links, established between the client and data servers, to transfer data blocks in parallel. The client is responsible for joining the downloaded blocks. The system and communication link capacity of the client can therefore be fully utilized. To enable highly dependable service, our method uses coding techniques to add redundancy to the original data. This redundancy enables us to retrieve the original data even if portion of the data is unavailable due to server and/or network failure. Our approach is much more cost-effective than replication because the encoded redundancy is much smaller in size than the original data. Unlike replication, load balancing is handled and

guaranteed by the system. Our stripping approach enables the client to fully utilize its bandwidth. The client bandwidth can be saturated by the aggregate bandwidth of multiple parallel data servers. Therefore, the server speed and bandwidth are never the bottleneck. What the client pays is what he gets. Throughout the project research, we will investigate theoretical and technological issues such as application-specific multimedia data block allocation and layout strategies, asynchronous block pre-fetching at each storage server, communication protocols (client/server, server/server), and dynamic evolution of reliability schemes.

The objectives for the proposed research are to:

- explore the characteristics of end-user information retrieval, organization and navigation in a digital library environment;
- explore ways in which conventional and advanced techniques for these processes could be applied in this environment;
- determine ways in which current techniques for these processes are inadequate or in need of improvement;
- develop new algorithms and processes that would point to the enhancement of the information storage, retrieval, organization, navigation and visualization processes.

## IX. FUTURE WORK

Following the Image Processing results, major questions arose from this year research. Some of them are:

- What are the effects of compression in the digital library – distortion, noise?
- Which is the most effective methodology to measure the memory required to store a file after applying the iterative technique?
- What is the optimum criterion to select the wavelet?
- Which are the adequate parameters to measure image quality?

Two different approaches can be taken to get images with different resolutions

1. Apply different wavelets for the decomposition process. However, there needs to be a relationship among those wavelets, maybe all from the same family, so the reconstruction of the image is only one procedure. A strong research using wavelet packets is recommended.
2. Using a certain wavelet, apply the decomposition process for different levels and study the coefficients so that an optimized criteria selects the coefficients to be manipulated.

In terms of the noise effects, two tasks are necessary to understand.

1. A clear visualization of coefficients behavior for a clean and noise image.
2. Application of lossless and lossy compression techniques to transmit images. Compare the results to analyze its effects. Take into consideration that two types of noise can be found, one due to compression and the other due to transmission.

A new area of research that will be proposed as a part of future research is that of novel computing architectures and the associated memory issues for teraflop and petaflops computing. Databases are the heart of many applications today and they will take an important role in future customer services.

## X. REFERENCES

1. HPCC-FY 1995 Implementation Plan, National Coordination Office for HPCC, Office of Science and Technology Policy.
2. Evolving the HPCC Initiative to Support the Nation's Information Infrastructure National Research Council, National Academy Press, 1995.
3. A. Linke et al., "A Tool for Materials Exploration " International Conference on high Performance Computing and Networking, Lecture Notes in Computer Science, vol. 919, p. 928, B. Hertzberger and G. Serazzi (Eds.), 1995.
4. O. Sharp, "the Grand Challenges, " BYTE, pg.65, February 1995.
5. "IMA to focus on Materials Science in 1995-96, "SIAM News, pg.18, Aug/Sept 1995.
6. B.D.Nordwall, "High Performance Software lags Behind Hardware," Aviation Week and Space Technology, pg.54, May 23, 1994.
7. F. Weingarten, "HPCC: Still a Percheron Horse," Communications of the ACM, vol. 38, pg.21, June 1995.
8. F. Weingarten, "HPCC Research Questioned," Communications of the ACM, vol. 36, pg. 27, November 1993.
9. I.M.Kameny, "Database Technology Activities and Assessment for Defense Modeling, and Simulation Office," Rand Corporation Report, 1994.
10. A.A. Kumar, "Data Management and Database Creation: Applications to Scientific and Engineering Database," Proc. NOBCCHE Conference, 1995.
11. D.Barry, "Co-Existence With Existing Systems, " Object Magazine, pg.78, June 1996.
12. T.Sterling et al., "Enabling Technologies for Petaflops Computing," MIT Press, 1995.
13. A.A.Kumar, "SEA HPCC Program- A New Paradigm for Twenty- First Century Computing," Presentation to the HPCC Design Team, Jackson State University, June 1996.
14. Mayra Caceres, Comparative Analysis of Image Compression Techniques, E&A "97 Proceedings, Vol. 1.
15. Anowal Huq, Simulation of Non-Gaussian Communication System, E&A "97 Proceedings, Vol. 1.
16. Andrew Cray and Peter Heywood, The New Public Network, Special Report, Data Communications, October, 1998.
17. Andreas Bittorf, How to set up an image database on the WWW efficiently, [http://matrix.ucdavis.edu/mirror/personen/asbittor/tut/imaag\\_db.htm](http://matrix.ucdavis.edu/mirror/personen/asbittor/tut/imaag_db.htm), 09/96.
18. G. Thoma and L. R. Long, Compressing and Transmitting Visible Human Images, IEEE Multimedia, April-June 1997.
19. David Doermann, The Indexing and Retrieval of Document Images: A Survey, Computer vision and Image Understanding, Vol 70, No. 3 , pp.287-298, 1998
20. G. Cha and C. Chung, A New Indexing scheme for Content-Based Image Retrieval, Multimedia Tools and Applications, Vol. 6, pp. 263-288, 1998.
21. M. Barazande-Pour, L.W. Mark, and A. K. Khaandani, Multi-level Priority Transmission of Images over a Turbo-Coded Channel, 1997 IEEE.
22. Susan Gauch, ProFusion Information, [www.ProFusion.com/info.html](http://www.ProFusion.com/info.html)

## APPENDIX A

### SELECTED RESULTS FROM THE WAVELET COMPUTATIONS

Testing the iterative algorithm with different wavelets, certain threshold (to visualize the effect on the number of coefficients), fixed number of levels for image decomposition and several iterations to obtain results in terms of percentage of zero coefficients and the retained energy are as follows. The image used to do the testing was Barbara - black and white image. Table 6 shows some of these results and figure 4 has an image example from this table.

**Table 6. First simulations using the iterative algorithm.**

# levels	Threshold	Wavelet	Iteration	% zeros	Energy	MSE
2	Nominal	db7	1	46.42	99.99	1.73
			2	46.42	100	0.0114
			3	46.42	100	0.0073
		db3.5	1	46.49	99.98	2.21
			2	46.49	100	0.01
			3	46.49	100	0.01
		coif2	1	46.49	99.98	1.56
			2	46.49	100	0.005
			3	46.49	100	0.003
		sym7	1	46.46	99.98	1.65
			2	46.46	100	0.01
			3	46.46	100	0.004
	2*nominal	db7	1	65.94	99.91	7.91
			2	53.23	100	0.003
			3	53.02	100	0.001
		db3.5	1	66.96	99.89	9.27
			2	57.67	100	9.27
			3	56.99	100	9.27

The next simulation was run selecting different global thresholds. This experiment selected 0.5 nominal, nominal and 6\*nominal threshold. The goal of this experiment was to identify the effect of eliminating coefficients. Table 7 shows the results and figure 5 has an example.

**Table 7. Simulations with drastic changes in the global threshold value.**

# levels	Threshold	Wavelet	Iteration	% zeros	Energy	MSE	
2	Nominal/2	bior3.5	1	27.94	99.99	2.53	
			2	28.01	100	0.0223	
			3	28.03	100	0.0	
	Nominal		1	46.49	99.91	17.46	
			2	46.49	99.99	0.0615	
			3	46.49	99.99	0.0296	
	6*nominal		1	90.31	97.84	355.83	
			2	85.24	100	0.0	
			3	84.90	100	0.0	
	Nominal/2	db7	1	28.25	100	1.5115	
			2	28.32	100	0.0147	
			3	28.37	100	0.0135	
			nominal	1	46.46	99.93	10.49
				2	46.46	100	0.0562
				3	46.46	100	0.0375
			6*nominal	1	88.91	98.40	255.17
				2	69.34	100	0.0
				3	67.62	100	0.0

A different set of simulations was performed with similar settings as those for table 7 except that here we used three levels of decomposition. These results are shown in table 8 with a related picture shown in figure 6.

**Table 8. Summary of results for three levels of image decomposition.**

# levels	Threshold	Wavelet	Iteration	% zeros	Energy	MSE
3	Nominal*2	Bior3.5	1	48.8889	99.9692	2.5718
			2	48.8884	99.9969	0.0161
			3	48.8884	99.9972	0.0166
		Coif2	1	48.8884	99.9778	1.9227
			2	48.8884	99.9985	0.0093
			3	48.8884	99.9986	0.0061
		Symb7	1	48.8071	99.9775	1.9126
			2	48.8071	99.9931	0.0133
			3	48.8071	99.9983	0.0081
	Nominal*4	Bior3.5	1	84.0251	99.4404	38.0129
			2	75.9699	100	2.53E-4
			3	74.8218	100	7.95E-7
		Coif2	1	82.8969	99.6356	31.5887
			2	57.3590	100	2.24E-5
			3	56.7498	100	4.8E-8
		Symb7	1	83.2442	99.6445	30.7663
			2	56.5132	100	2.39E-4
			3	55.3883	100	3.13E-6

Five levels of decomposition and non-uniform changes in the threshold value were also performed. The simulation changed the previous values and kept the same wavelet. These results can be seen in Table 9 and a picture can be seen in figure 7.



**Table 9. Non-uniform changes in the threshold value for each iteration.**

# levels	Threshold	Wavelet	Iteration	% zeros	Energy	MSE
5	[2 4 6 8 10]	Db7	1	74.2244	99.0144	32.1049
			2	57.6203	99.9980	0.0213
			3	56.8642	99.9986	0.0029
	[4 5 5 6 6]		1	84.7935	99.5060	42.2976
			2	61.5319	99.9983	0.0105
			3	61.0480	99.9988	0.0037
	[5 5 6 6 7]		1	87.2924	99.3893	52.8321
			2	60.8148	99.9981	0.0114
			3	60.3599	99.9987	0.0032

Another simulation was performed changing the threshold value for the vertical (v), diagonal (d) and horizontal (h) details. Let us keep in mind that the goal of these simulations is to visualize the effects of eliminating coefficient to generalize the criteria for optimization. This setting has five levels of decomposition, the db7 wavelet and it will analyze the percentage of zeros and the energy of the image. Table 10 has some of these results and a set of very interested pictures from this table are shown in figure 8 (without vertical details), figure 9 ( without horizontal), and figure 10 (without diagonal details).

**Table 10. Effects of the threshold on each detail.**

# levels	Threshold	Wavelet	Iteration	% zeros	Energy	MSE
5	2 h	Db7	1	54.90	99.76	37.61
	0.5 d		2	45.55	100	0.0114
	2 v		3	45.44	100	0.0046
	2 h		1	76.49	93.23	125.4
	25 d		2	62.71	100	0.0
	2 v		3	62.17	100	0.0
	2 h		1	78.21	97.14	452.08
	2 d		2	62.71	100	0.0
	25v		3	61.81	100	0.0
	25 h		1	76.77	99.14	120.14
	2 d		2	61.83	100	0.0
	2 v		3	60.97	100	0.0

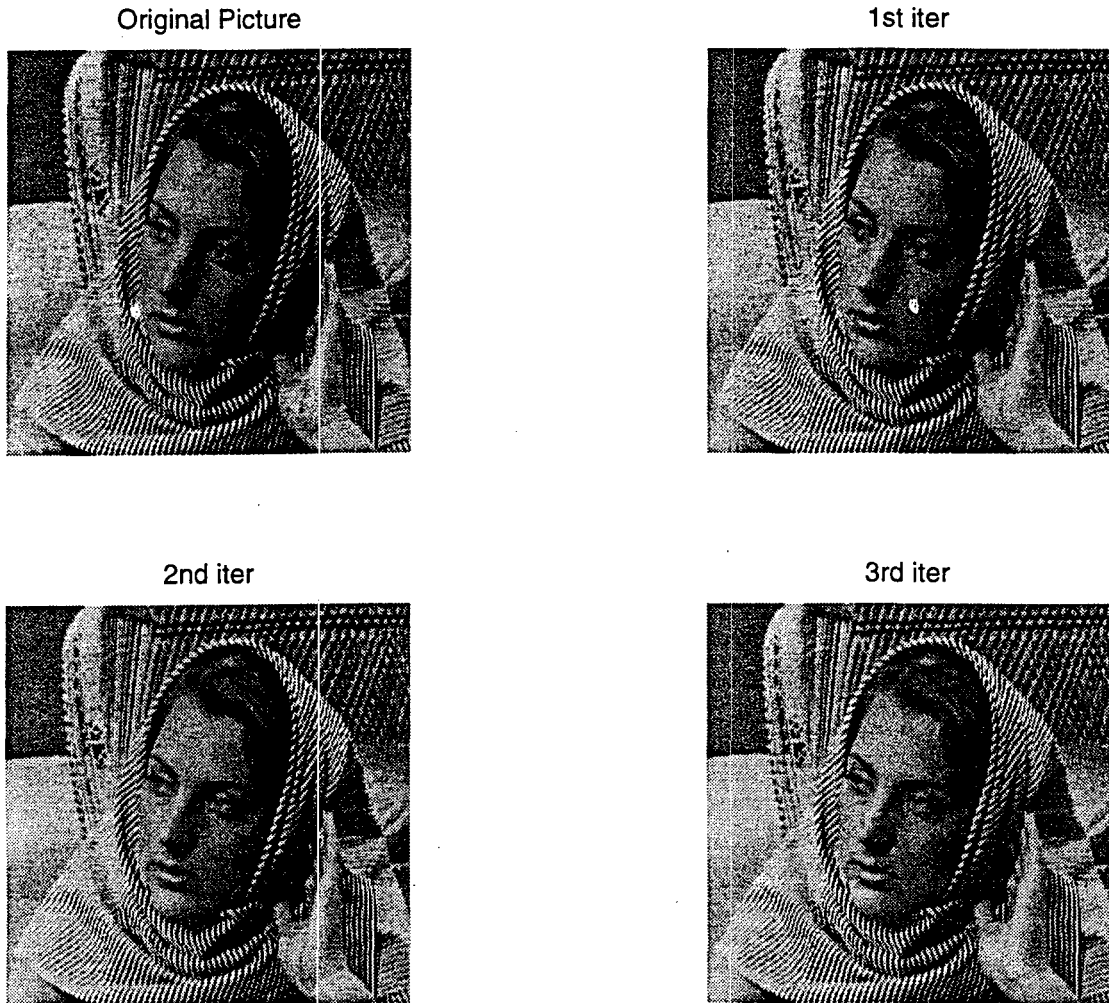
Also, simulations were performed using color images with the same settings used for table 10. The following pictures are some of the results gathered from the previous tables. For more detail information there is a separate report available upon request.

The last set of pictures (figure 11), represent a progressive recovery or retrieval, from low to higher resolution until the reconstruction of the original image. This simulation was performed without eliminating any coefficient that means without threshold. However, more work needs to be done in this area so have a solid criterion of determining the coefficients to be eliminated.

## Conclusions

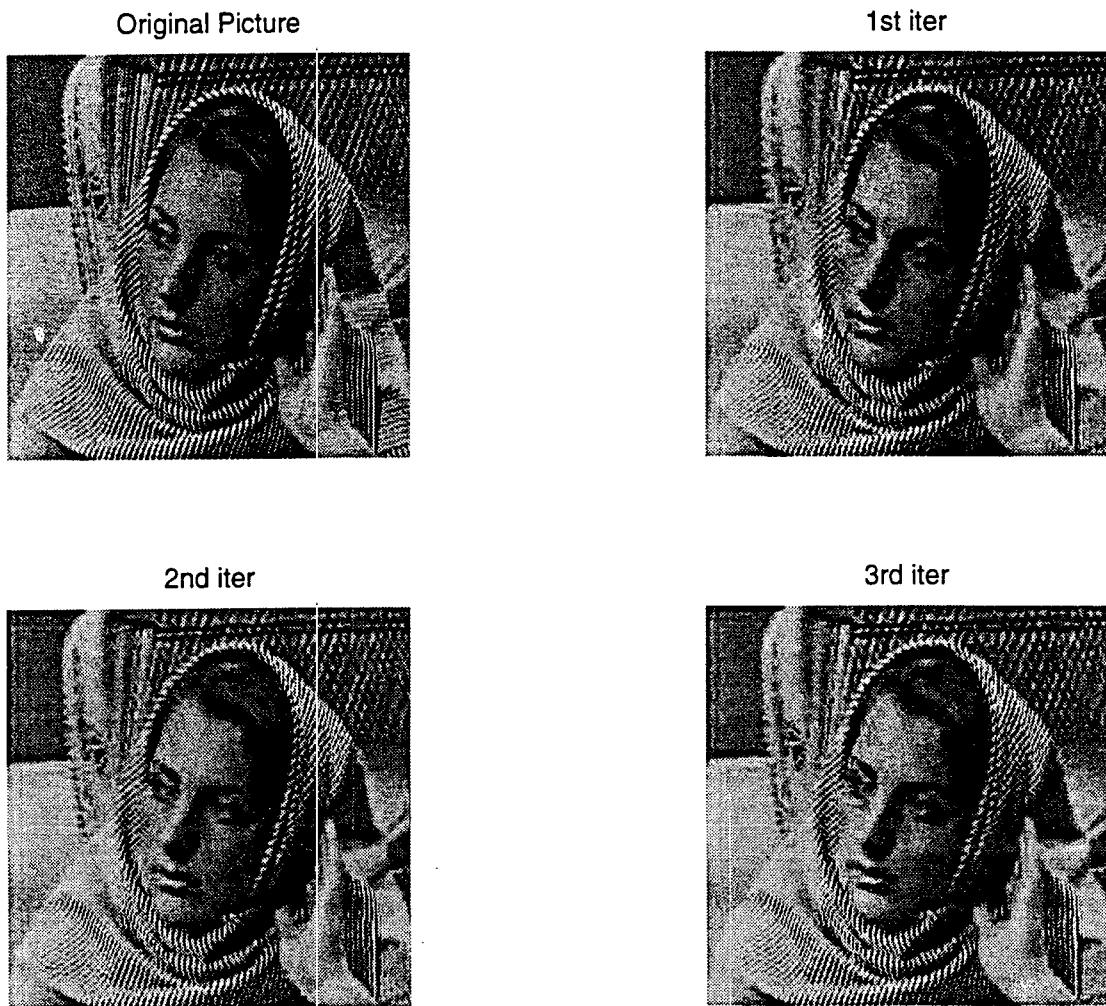
1. A common situation for all the simulations is the fact that for the third iteration the number of zeros as well as the energy of the image does not differ that much from the values for the second iteration. This is the idea behind the “Fixed Point”.
2. Progressive retrieval was obtained based on the levels of decomposition. However, a complete study on determining the best way to select the coefficients to be transmitted need to be done in future research.
3. The criterion to eliminate coefficients is not very clear. Many different effects on the images based on the changes of the threshold values that can not allow establishing a pattern. More work needs to be done in this topic.
4. The number of level for decomposition of the image seems to be irrelevant. This will decrease the calculation time of the algorithm.
5. The selection of the wavelet for each image was done randomly and the results did not show a particular effect on the resulting picture.
6. There is the need for optimum measures of image quality as well as the understanding of the effect of noise due to compression.

**Figure 4 – Simulation using the iterative algorithm**



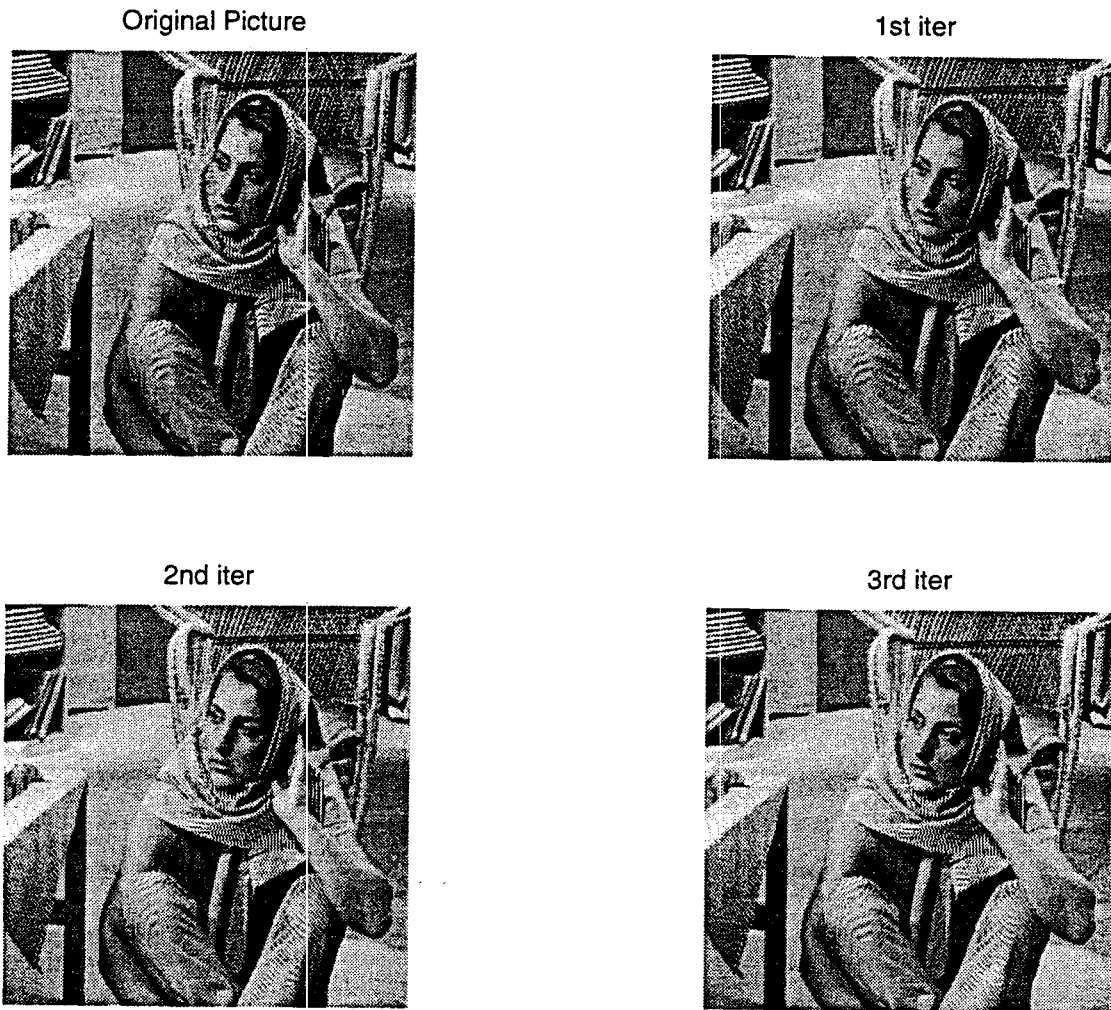
**This example shows the image called wbarb for three different iterations. The settings are number of levels equal 2, the wavelet is bior3.5, the threshold is the nominal threshold. As you can see from these pictures and the values on table 6, the results for iteration two and three are very much the same.**

**Figure 5 – Simulation using drastic changes on the global threshold**



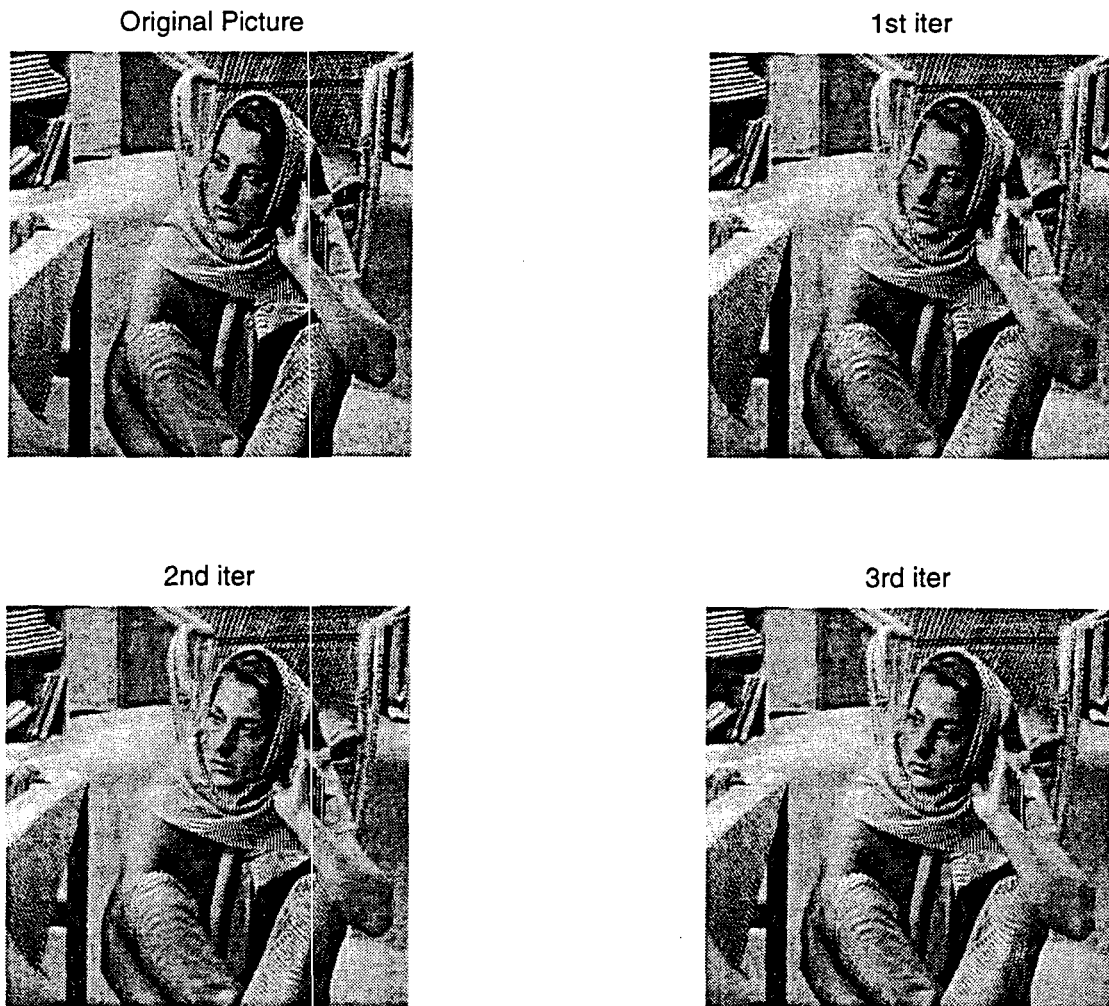
This example shows wbarb for three different iterations. The settings are number of levels equal 2, the wavelet is db7, the threshold is the  $6 \times$  nominal threshold. As you can see from these pictures and the values on table 7, the results for iteration two and three are very much the same. This is valuable for getting the Fixed point image.

**Figure 6 – Simulation using more levels of decomposition**



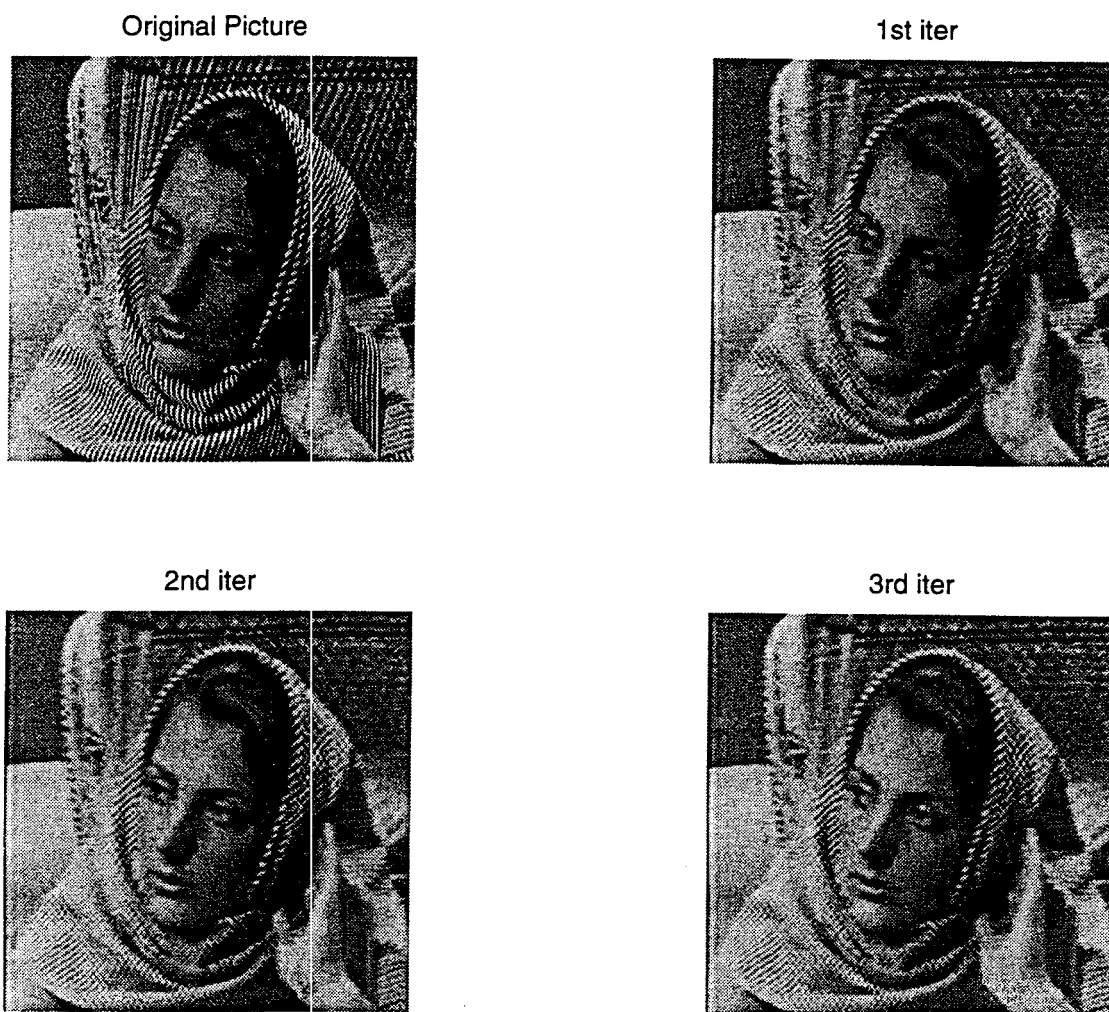
This example shows wbarb for three different iterations. The settings for this simulation are number of levels equal 3, the wavelet is coif2, the threshold is  $2 \times \text{nominal threshold}$ . As you can see from these pictures and the values on table 8, the results for all iterations are very close to each other.

**Figure 7 – Simulation using non-uniform changes for the threshold in each iteration**



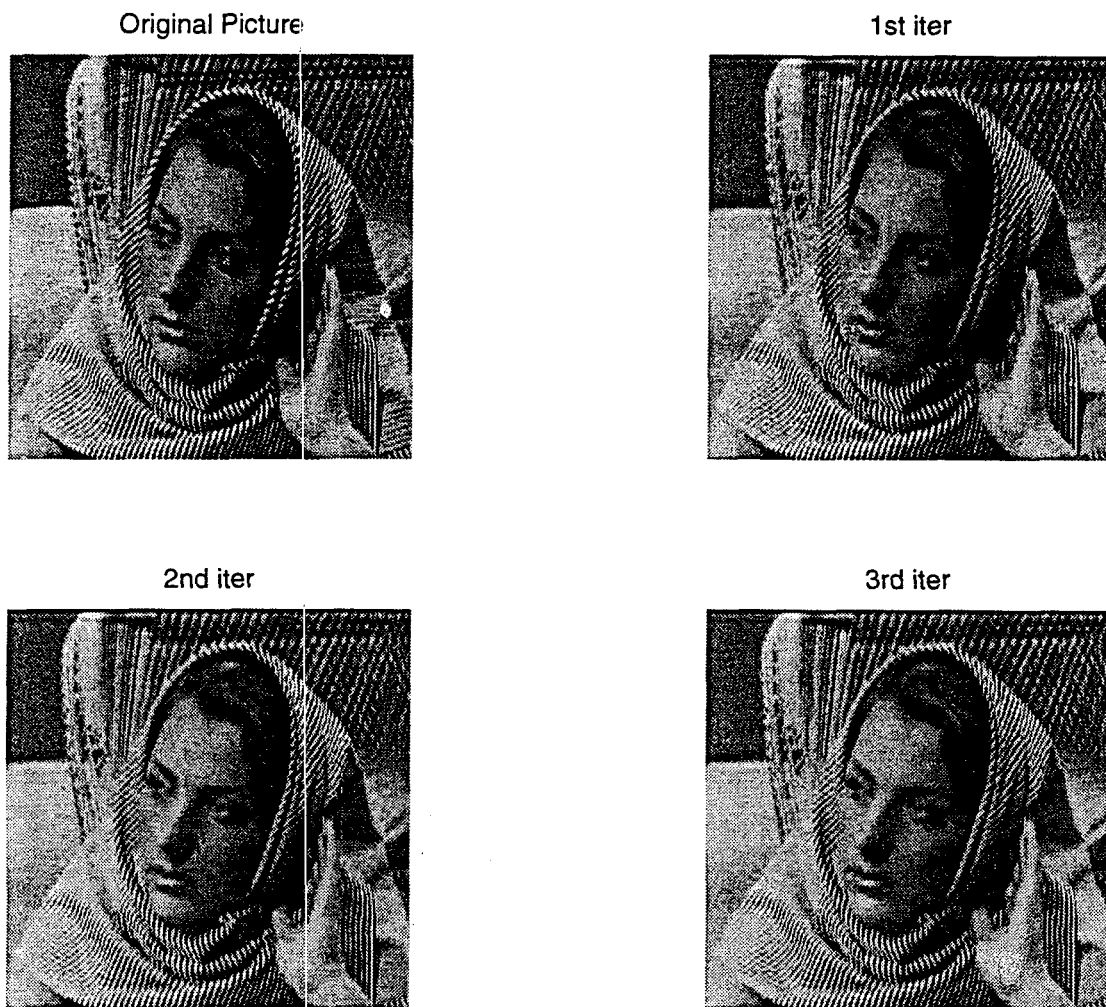
This example shows wbarb for three different iterations. The settings for this simulation are number of levels equal 5, the wavelet is db7, the threshold is equal to the following values for each iteration,  $\text{thr} = [2\ 4\ 6\ 8\ 10]$ . As you can see from these pictures and the values on table 9, the results for all iterations are close but you can see differences on the pictures.

**Figure 8 – Simulation using drastic changes for the threshold in each detail.**



This example shows wbarb for three different iterations. The settings for this simulation are number of levels equal 5, the wavelet is db7, the threshold is equal to the following values for each detail,  $\text{thr} = [2 \ 2 \ 25]$  for horizontal, diagonal and vertical details. As you can see from these pictures and the values on table 10, the results for all iterations are close but you can see differences on the pictures.

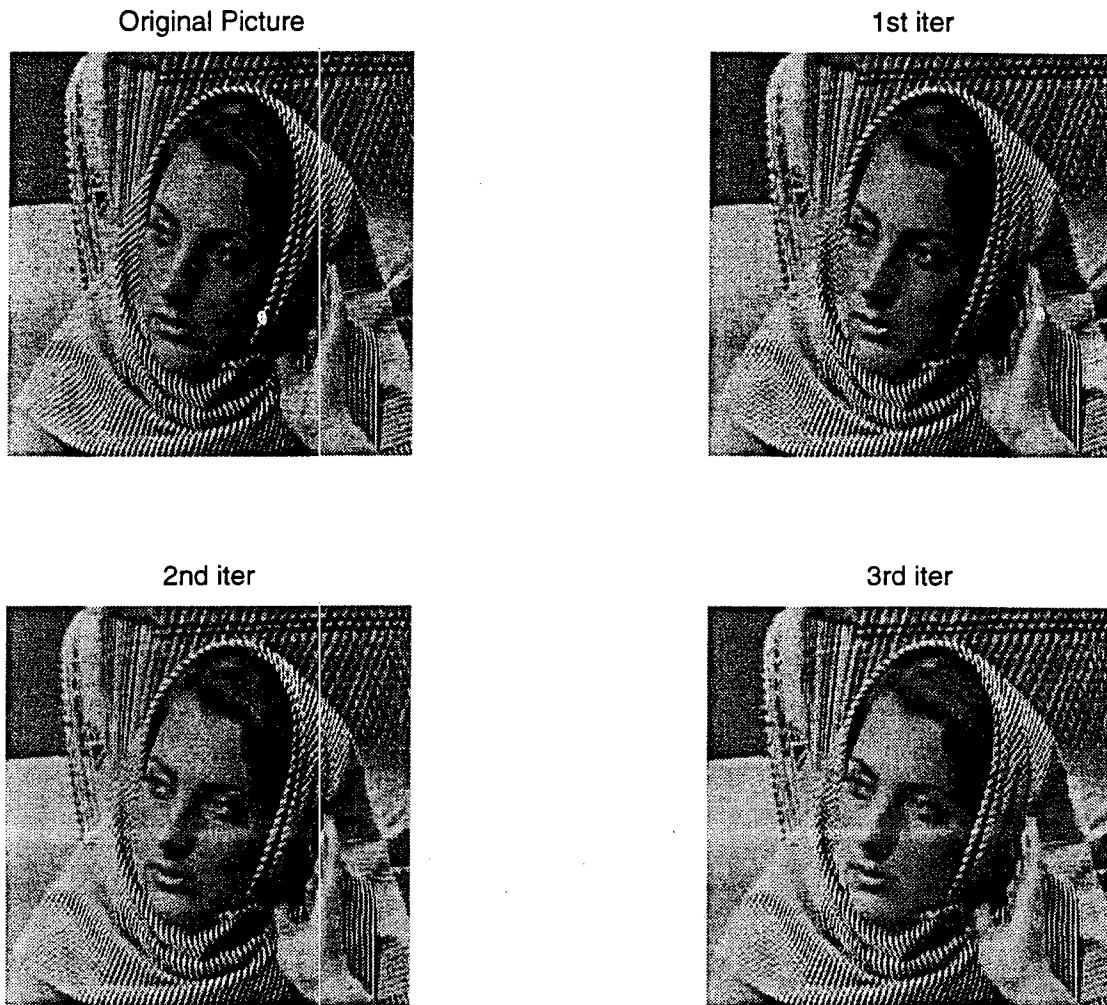
**Figure 9 – Simulation using drastic changes for the threshold in each detail, here we eliminate horizontal details.**



This example shows  $wbarb$  for three different iterations. The settings for this simulation are number of levels equal 5, the wavelet is db7, the threshold is equal to the following values for each detail,  $thr = [25\ 2\ 2]$  for horizontal, diagonal and vertical details. As you can see from these pictures and the values on table 10, the results for all iterations are close but you can see differences on the pictures.

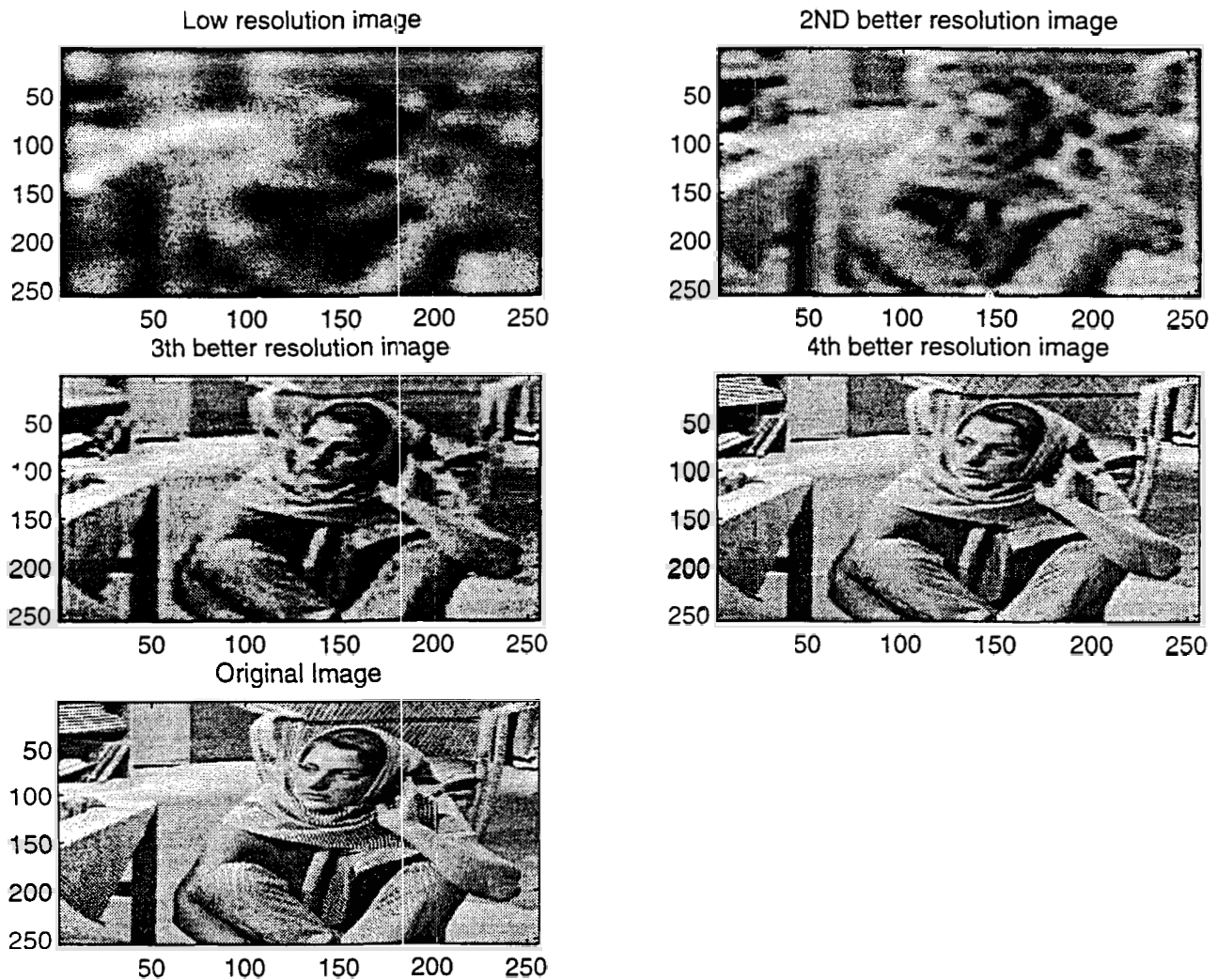


**Figure 10 – Simulation using drastic changes for the threshold in each detail eliminating the diagonal details.**



This example shows wbarb for three different iterations. The settings for this simulation are number of levels equal 5, the wavelet is db7, the threshold is equal to the following values for each detail,  $\text{thr} = [2 \ 25 \ 2]$  for horizontal, diagonal and vertical details. As you can see from these pictures and the values on table 10, the results for all iterations are close but you can see differences on the pictures.

**Figure 11 – Simulation for progressive retrieval based on resolution.**



This example shows the results for different resolutions. The settings for this simulation are number of levels equal 4, the wavelet is db7, the threshold is equal zero therefore there is no elimination of coefficients. As you can see from these pictures the third iteration is a good start (2<sup>nd</sup> better resolution image) to reconstruct the original image if the user requires better image quality.

## APPENDIX B

### Image Processing Techniques for Digital Image Libraries Mayra Caceres AT&T Holmdel, NJ

The major goals of this thesis were to examine different techniques of image compression, in particular, to introduce a new techniques based on wavelet transform. As can be seen in the book, these goals have been attained. A summary of important issues on the image compression techniques implemented in this thesis.

There are many techniques for image coding and compression. Transform coding based on the Discrete Cosine Transform (DCT) is popular since the 1980s because of low complexity and effective bit allocation. This became the JPEG standard in image coding. For gray-scale images, it performs well for compression ratios up to 16 to 1. At 24 to 1 compression, JPEG's synthesized image suffers from blocking effects that cannot be accepted, which manifests the short basis functions used in reconstruction.

Subband coding is very successful. Pyramid coding was and is effective for high bit-rate compression. Subband coding using wavelets (this means tree-structured filter banks) avoids blocking at medium bit-rate, because its basis functions have variable length. Long basis functions represent flat background or low frequency. Short basis functions represent regions with texture. At a reasonable distance, one can not detect the errors easily. At low bit-rate, wavelet coding suffers from ringing when high frequencies (textures) are deleted. The ringing artifacts are significant around edges with high intensity. They have the shape of the basis functions that are emphasized in synthesis.

The DCT preserves energy and the essential information is in the first few coefficients. Quantization assigns more bits to these pixels. The objective of bit allocation is to minimize the distortion. The quantized subbands are then scanned and coded using lossless compression. This entropy coder watches for runs of zeros and transmits their length (roughly 3:1 for free). The wavelet-based transform coder can be seen in similar way. For one-level decomposition, the approximation image is obtained by low-pass filtering in both the horizontal and vertical directions (LL). The other three sub-images have details involving high frequencies, being the horizontal details the result of high-pass and low-pass filtering (HL), the vertical details the result of low-pass and high-pass filtering (LH), and the diagonal details the result of high-pass filtering in both directions (HH). The bit allocation algorithm will assign many bits to LL and few bits to HH. The normal number of iterations on the LL sub-images is 4 or 5.

Because of the superior energy compaction and correspondence with the human visual system, wavelet compression has produced good results. Since the wavelet basis functions have short support for high frequencies and long support for low frequencies, large smooth areas of an image may be represented with very few bits. High frequency detail is added where it is needed. Huffman coding is used in JPEG algorithm, as well as other compression schemes. The Huffman code assignments produce what is called a minimal code. This is optimal encoding, when creating a one-to-one coding scheme. Huffman codes typically produce compression ratios in the range of 1.5:1 to 2:1, but the exact amount of compression is data set dependent.

People working with wavelets think that it should be developed to be a wavelet image compression standard more general than just for fingerprints. Also working with moving pictures, make wavelet methods work for video compression. While MPEG's success and proliferation make medium bit-rate wavelet video a purely academic endeavor, low bit-rate compression remains a place where wavelet methods can prevail.

At high and medium bit rate (low and medium compression ratio), there is a strong correlation between the image quality and these objective measures. A good compression algorithm would reconstruct the image with low MSE and high PSNR. At low bit rate (high compression), the basis function characteristics have profound effects on the quality of the reconstructed image.

### **Estimation of the Channel Transfer Function Using Higher Order Statistics** **Anowarul Huq, Caterpillar Corp., Chicago, Illinois**

Signal processing with Higher Order Statistics has not yet become a full-blown subject, but it does carry a substantial future especially in the creation of advanced communication, Sonar, Radar, Speech, Geophysical and image processing area. There is a wide variety of operations can be done with the help of higher order statistics and particularly with higher order spectra. An attempt is made here to choose one of the processes, bispectrum, and have tried to explore the subject. The goal, the estimation of the channel transfer function, is accomplished by reconstructing the original signal. The usefulness of the simulation program is as follows: 1) it is a fully functional program, that is for a different set of input data one can use this program to estimate the channel transfer function of a system and then can reconstruct the signal. In future, one may work on trispectrum (Fourier transform of the fourth order cumulant) the next step to bispectrum. The future work could be accomplished by modifying the simulation program to include the trispectrum process. By incorporating higher order spectra in the program, one can evaluate the performance of the added technique in the reconstruction of the input signal. After evaluating the performance of the process bispectrum and that of trispectrum, one can decide whether it is necessary to include trispectrum or other higher order spectrum in order to process the signal. In addition to the reconstruction of the signal using bispectrum, this program can also be used to process for the detection and characterization of non-linearities in time series of signals and systems.

**Section 4. High Performance Computing in High- $T_c$   
Superconductivity**  
**J. D. Fan**  
**Southern University and A&M College**

# **Southern University and A & M College**

Department of Physics  
P. O. Box 10554  
Baton Rouge, LA 70813

E-mail: fan@stark.phys.subr.edu

Tel & Fax: (504) 771-3926

## **Annual Progress Report to Doe through SEA**

for the Subcontract to the Grant No. DE-FG02-94ER25229

By

J. D. Fan, Ph.D. Principal Investigator  
Department of Physics, Southern University  
Baton Rouge, LA 70813

Sep. 15, 1997-Sep. 14, 1998

The project funded by DOE under the grant captioned above the SEA, Inc. was extended for one year beyond the original length of time without any additional fund, provides us an opportunity to continue our efforts in superconductivity carried out during the past years. As proposed, we planned to do both simulational and computational study of superconductor structures and properties. In order to proceed them, we have to simultaneously undertake analytical calculations that serve as a foundation of the computing work and simulational study.

### **1. Simulational and Computational Studies**

1) Two simulational studies of high-temperature superconductor structures have been finished and papers published.

2) Two ab initio calculations of electronic structures have been completed and papers published.

3) MD simulation study of temperature dependence of defects of high-temperature superconductors is being performed as a MS thesis. It is basically completed and the thesis is being prepared by the student.

4) We have encountered a difficulty in studying simulational calculations for two types of atoms in a system. The configurations on

computer cannot reach equilibrium. It seems that more analytical work is needed and C-codes may also possess errors.

Due to the outage of fund the research has to be stopped.

### Scientific Activity

This year, we have engaged in a public scientific activity by organizing an international conference on superconductors and related materials in Baton Rouge. More than 150 attendees have participated in this conference and an anticipated success has been reached.

## 4. Professional Presentations

During the period from September, 1997, through September, 1998, the total number of professional presentations that were made is **sixteen (16)**, out of which **seven (7)** were made internationally or abroad.

International and Abroad	National	Local	Total
7	7	2	16

## 5. Professional Publications

- 1) J. D. Fan, G. L. Zhao, T. Edis and Y. M. Malozovsky, "A Molecular Dynamics Simulation Study of the Two-Dimensional Lattice Structure of the Cu-1 Plane in YBCO." Phys. Rev. B, **56**, 10747 (1997).
- 2) T. Edis, J. D. Fan, D. Bagayoko, G. L. Zhao and J. T. Wang, "A Molecular Dynamics Simulation of the Two-Dimensional Cu-O Chain Structure in  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ ," IJMPB, Vol. I&II, Dec. (1998) in press.
- 3) D. Bagayoko, G. L. Zhao, J. D. Fan and J. T. Wang, "ab initio calculations of the electronic structure and optical properties of ferroelectric teragonal  $\text{BaTiO}_3$ ." J. Physics: Condensed Matter **10** (1998) 5645-5655.
- 4) Bagayoko. G. L. Zhao, J. D. Fan and J. T. Wang, "Half-Mettalic Behavior in  $\text{La}_{33/4}\text{Ca}_{1/4}\text{MnO}_3$ ," IJMPB, Vol. I&II, Dec. (1998) in press.
- 5) J. D. Fan and Y. M. Malozovsky, "A Pairing Mechanism in Superconductivity due to the screened Coulomb Field," *Superlattices and Microstructures*, Vol. 21, No. 2 (19097) 209-215.
- 6) J. D. Fan and Y. M. Malozovsky, "Competition Between the Coulomb and Phonon Fields: Superconductivity in a Quiescent Fermi Gas," *Superlattices and Microstructures*, Vol. 21, No. 3 (1997) 299-303.

- 7) Y. M. Malozovsky and J. D. Fan “Cooper Instability for a Given Field in a quiescent Fermi Gas,” Intl. J. Mod. Phys. B. Vol. 12, No. 6 (march 10, 1998) pp. 637-652.
- 8) J. D. Fan and Y. M. Malozovsky, The role of the Coulomb Interaction in Superconductivity,” IJMPB, (Vol. I&II, Dec. (1998) in press.
- 9) Y. M. Malozovsky and J. D. Fan, “Impurity-Doping in High-Tc Superconductors: Anderson Localization vs. Interaction Effect,” J of Korean Physical Society, Vol. 33, No. 5, 589--593 (1998)



**Section 5. Science and Engineering Alliance**  
**Dr. Robert Shepard**

# **Subcontractor Final Report**

DE-FG02-94ER25529

Project Period 09/15/94-09/14/98

**SEA: Building Partnerships for the Future**

A Report on the  
SEA High Performance Computing and Communication (HPCC)  
Initiative

by

Robert L. Sthepard, Ph.D.  
Executive Director, SEA  
November 1998

## **I. Introduction**

### **A. Purpose of Report**

This document reports on the accomplishments and beneficial support emanating from the relationship between the Office of Executive Director (OED), Science and Engineering Alliance (SEA), and the alliance member institutions in establishing the SEA high performance computing and communication (HPCC) initiative.

### **B. Background**

Since 1990, the SEA has worked diligently in its quest to create a model partnership involving four state-supported Historically Black Colleges and Universities (HBCUs) and a national laboratory. This relationship has expended scientific and engineering opportunities to the four alliance members HBCUs. The partnership between the academic institutions, i.e., Alabama A&M University, Jackson State University, Prairie View A&M University, Southern University and A&M College, and national laboratory, Lawrence Livermore national laboratory, has seen much growth in the years since its inception. The SEA partnership was recognized as a competent technical resource by the U. S.

Department of Energy's (DOE) Office of Scientific Computing. Jackson State University was prime contractor on the HPCC initiative, with the other SEA members, including the OED, serving as subcontractors.

The SEA has successfully taken on multiple ventures, and their associated responsibilities, to continue the exposure of the alliance institutions. Each of these ventures have been beneficial to enhancing the total infrastructure of the academic institutions, i.e., students faculty, administrative staff and physical facilities.

## **Subcontractor Final Report**

DE-FG02-94ER25229

Project Period: 09/15/94-09/14/98

### **II. Role of the Office of Executive Director (OED) in Establishing SEA High Performance Computing and Communications (HPCC) Initiative**

#### **A. Network Developer (people first)**

Networks connect qualified, competent people together for a common purpose. In the case of the SEA HPCC initiative, these "connections" served as points of access into areas of scientific and engineering endeavors to fully integrate qualified and competent HBCU faculty and students into mainstream HPCC technical research activities.

In February 1993, the OED held a planning meeting on the campus of Jackson State University. The purpose of this meeting was to develop a network of SEA technical expertise for moving forward with an expanded collaborative initiative in the area of HPCC. The meeting resulted in the formation of an SEA-HPCC Design Team (SDT).

The role of the SDT was to develop a document describing the collective capability of the SEA in the area of HPCC-related activities based on the long-term direction of DOE's HPCC initiative. The SDT developed a criteria for soliciting abstracts relating to HPCC activities from a wide cross section of SEA researchers. By March 1993, the SDT received the abstracts. By May 1993, the SDT screened the collection of abstracts and developed a "Talking Paper for Marketing Purposes (TPMP)." The TPMP was composed of an executive summary, expertise tables, and was classified into six subareas of research:

- Global Change;
- Biotechnology and Biomedical;
- Data Acquisition, Storage, Processing and Integration;
- Toxic Waste;
- High Performance Computational Issues in Materials Science and High Energy Physics, and
- Outreach.

#### **B. Market Representative**

In June 1993, the OED launched efforts to market the TPMP. The TPMP reflected the SEA member institutions' existing collective strengths in the area of HPCC. In January 1994, the DOE showed interest and chose one project from each institution as possible candidates for funding. The selected candidates from the TPMP were further developed into an unsolicited proposal.

In April 1994, the SEA-HPCC proposal was officially submitted to the DOE. Following peer review and modifications, the SEA-HPCC proposal was funded in FY 1995 (late calendar year 1994) by DOE's Office of Scientific Computing for three years at a revised funding level of \$1,069,200. The HPCC initiative became the SEA's first multiyear grant.

Much of the next phase of marketing involved the OED creating opportunities to familiarize the broader technical community concerning the capabilities of HBCUs to conduct investigations in the area of HPCC.

### **C. Facilitator**

Following full implementation of the SEA-HPCC initiative at the campus level of the alliance member institutions, the OED moved to expand the initiative by creating organizational networks and connections with public and private organizations and national and international groups. Thus, establishing bridges for the alliance institutions' through national conferences and meetings with other technical groups, because the primary strategy for expanding the SEA-HPCC initiative. The efforts of the OED resulted in the following:

- 1994 --- CRAY Research, Inc. makes an offer to expand the SEA-HPCC initiative.
- 1995 -- The SEA-HPCC PIs hosts a forum at the 22nd national Conference of the National Organization of Black Chemist and Chemical Engineers (NOBCChe). The SEA executive director served as Session Chair for the forum that allowed him to acquaint the conferees with the work, experience and Capabilities of the alliance institutions.
- 1996 -- Strategic meeting arranged between the SEA-HPCC PIs and the Lawrence Berkeley national laboratory (LBNL) to forge a partnership between the two entities. The meeting identified several areas showing potential for collaboration between LBNL and the SEA.

- 1997 -- Research Consortium Innovative (RCI), Ltd. invites OED to participate in discussions to bring SEA-HPCC program into mainstream national and international Supercomputing arena.
- 1998 -- SEA Member, Southern University, Selected to Host Gulf Coast Regional Climate Change Workshop. Visit web site: [www.usgcrp.gov/regional.html](http://www.usgcrp.gov/regional.html).
- 1998 -- SEA, along with Cornell Theory Center, named Associate Member of RCI, Ltd. through the sponsorship of IBM Corporation. See 2nd Quarter, Fall 1998, under RCI web site: [www.rciltd.com/newsletterMain.htm](http://www.rciltd.com/newsletterMain.htm).

### **III. The Emerging Role of SEA**

While the SEA was born out of a reduction in the resources for education and the government and industry's commitment to improve science and mathematics education, it has emerged as a multi-faceted mentor and champion of the capabilities of each of its member institutions. SEA continues to build successful relationships between the member institutions and the national laboratory. It has also expanded opportunities for inclusion of member institutions in federal scientific research and development initiatives.

### **IV. Profile for the Future**

For the benefit of the SEA member institutions, the OED must continue serving as mentor, role model, challenger, coach, market representative, promoter, facilitator and consultant. Additionally, the OED must strive to help alliance member institutions and students prepare for the future by anticipating the opportunities of the next millennium.

The U.S. Department of Labor predicts dramatic changes in the national labor pool in the future. Different perspectives, orientations and insights will breathe new life into institutions of higher learning and public and private organizations to meet the challenges ahead, the OED role will be to educate organizations and firms to the benefits and advantages of tapping the non-traditional source of expertise and talent of alliance member institutions.

Some advantages to an organization that values partnering with non-traditional institutions and a diverse applicant pool are:

- Access to a wider talent pool.

- Opportunity to develop overlooked talent pool.
- More astute strategies for mission accomplishment coming from diverse perspectives.
- Innovative project, action, and performance planning can emerge from listening to points of view different from traditional sources.
- Better customer service can be offered by a diverse workforce that reflects and understands the different needs and preferences of those who make up the general population.

## **V. Conclusion**

The OED and alliance member institutions have much work ahead. There are many challenges and undiscovered territory, but the SEA has a record of accomplishment and a model that works.

The SEA-HPCC initiative, funded by DOE, has provided SEA researchers access and has fostered their extensive engagement into the broad field of HPCC. The initiative is enabling the alliance member institutions to make significant contributions to the full range of HPCC activities, including existing supercomputing systems, special purpose and experimental systems and the new generation of large scale parallel architecture.

The SEA marketing strategy takes many forms. The OED has been very skillful in exposing the alliance member institutions' researchers to the state-of-the-art in the area of supercomputing. Attendance and participation by SEA researchers in the National Supercomputing Forum identifies another access point for the alliance member institutions. SEA membership in RCI adds tremendous value to the acceptance and respectability of the investigations being conducted by SEA researchers.

A primary benefit of the Sea model is that it simultaneously provide institutional and career opportunities. Students who attend the alliance member institutions gain a competitive advantage by having research experiences and mentors and coaches in cutting-edge technical areas like HPCC. The fact that three African-American students obtained Master's degrees while working on the SEA-HPCC initiative supports these conclusions.

## **VI. Cost Data**



Cost Elements	Approved Budget
Fringe	19,476
	3,750
& Supplies	1,500
ons	2,000
e/Communication Services	2,340
Shipping	<u>634</u>
<b>Total for OED</b>	<b>\$29,700</b>

**Section 6. Project Administration**  
**Kunal Ghosh**  
**Jackson State University**

## **PROJECT ADMINISTRATION**

### **Final Report**

**The project administration was very heavily involved in all aspects of the sub-contracting process between Jackson State University and the other institutions and the SEA Executive Director's office. It routinely followed up on all communications between the institutions, irrespective of which office the original communication started. The facilitated to a great extent smooth and speedy starting of the different project components.**

**One of the primary roles of the project administration was to facilitate, smoothen, and accelerate the processes of the inter-institutional contracting, invoicing, and reimbursement. Also, the communication with the Department of Energy on different matters were funneled and accelerated. In case of Jackson State University research, all processes related to personnel, procurement, receiving, etc. were handled by the project office so that the PI could concentrate more on the research aspects.**

**The project administration consisted of the part-time project director, Dr. Kunal Ghosh, and a part-time administrative assistant, Ms. Debra Jackson. The project office was located in Rm. 202C/1, Just Hall of Science, on the Jackson State University campus.**